

The 1996 Census Unpaid Work Data Evaluation Study

Leroy O. Stone

and

Sandra Swain

Statistics Canada



Status of Women
Canada

Condition féminine
Canada

Canada

**The 1996 Census Unpaid Work Data
Evaluation Study**

Leroy O. Stone
and
Sandra Swain
Statistics Canada

The research and publication of this study were funded by Status of Women Canada's Policy Research Fund. This document expresses the views of the authors and does not necessarily represent the official policy of Status of Women Canada or the Government of Canada.

March 2000

Status of Women Canada is committed to ensuring that all research produced through the Policy Research Fund adheres to high methodological, ethical and professional standards. The research must also make a unique, value-added contribution to current policy debates, and be useful to policy makers, researchers, women's organizations, communities and others interested in the policy process. Each paper is anonymously reviewed by specialists in the field and comments are solicited on:

- the accuracy, completeness and timeliness of the information presented;
- the extent to which the analysis and recommendations are supported by the methodology used and the data collected; and
- the original contribution that the report would make to existing work on this subject, and its usefulness to equality-seeking organizations, advocacy communities, government policy makers, researchers and other target audiences.

Status of Women Canada thanks those who contributed to this peer review process.

Canadian Cataloguing in Publication Data

Stone, Leroy O., 1936-

The 1996 census unpaid work data evaluation study

Text in English and French on inverted pages.

Title on added t.p.: Étude d'évaluation des données du recensement de 1996 relatives au travail non rémunéré

Includes bibliographical references.

Issued also in electronic format through the Internet computer network.

ISBN 0-662-28018-0

Cat. no. SW21-40/2000E

1. Home economics — Canada — Statistics — Evaluation.
2. Women — Canada — Economic conditions — Statistics — Evaluation.
3. Sexual division of labour — Canada — Statistics — Evaluation.
4. Canada, Census, 1996.
5. Social surveys — Canada — Evaluation.

I. Swain, Sandra Leigh.
II. Canada. Status of Women Canada
III. Title.

HQ1381.5S76 2000 339.5'088'649'0971 C99-980275-5

Project Manager: Julie Dompierre, Status of Women Canada

Publishing Coordinator: Mary Trafford, Status of Women Canada

Editing: PMF Editorial Services Inc.

Translation: Maya Berbery

For more information contact:

Research Directorate

Status of Women Canada

350 Albert Street, 5th Floor

Ottawa, Ontario K1A 1C3

Telephone: (613) 995-7835

Faxsimile: (613) 957-3359

TDD: (613) 996-1322

E-mail: research@swc-cfc.gc.ca

This document is also available for download on the Status of Women Canada Web site at: <http://www.swc-cfc.gc.ca/>

TABLE OF CONTENTS

PREFACE	iii
ACKNOWLEDGMENTS	v
IMPORTANCE OF THE CENSUS UNPAID WORK DATA AND MAIN ISSUES ADDRESSED IN THE EVALUATION STUDY	
Introduction.....	1
Main Issues Addressed	1
Selected Findings.....	2
Major Observations on the Findings.....	3
PAPER 1: NON-RESPONSE AND IMPUTATION FOR MULTIPLE RESPONSES.....	
Introduction.....	5
Background	5
Data Quality	5
Non-Response Rate	6
Multiple Response Rate	9
Edit and Imputation	10
Imputation of Non-Response	11
Resolution of Erroneous Multiple Responses	12
Conclusion	12
Appendix 1.A: Census Unpaid Work Questions and Guide Instructions	14
PAPER 2: QUALITY OF THE 1996 CENSUS UNPAID WORK DATA IN LIGHT OF COMPARISONS WITH 1996 GSS DATA BASED ON IDENTICAL QUESTIONS	
Introduction.....	16
Methodology	16
Comparability of the Census and the GSS	17
Global Differences	20
Population Subgroups with Unusually High Levels of Census—GSS Divergence	24
Conclusion	29
PAPER 3: QUALITY OF THE 1996 CENSUS UNPAID WORK DATA IN LIGHT OF COMPARISONS WITH THE 1992 GSS BASED ON DIARY DATA	
Introduction.....	31
Methodology	31
Comparability of the Census and the GSS	31
Stylized versus Time-Diary Methods of Data Collection	32
Conclusion	38

Appendix 3.A: Operational Definitions of Destinations of Work of Economic Value	40
BIBLIOGRAPHY	41
GLOSSARY	42
ENDNOTES	43

PREFACE

The decision to produce the *1996 Census Unpaid Work Data Studies* originated from Status of Women Canada's discussions with the Policy Research Fund External Committee during the spring of 1998. The Committee identified unpaid care-giving work as an urgent policy research issue requiring immediate attention. Statistics Canada was approached to provide an early analysis of the data coming out of the 1996 Census of Canada which, for the first time, contained three questions concerning unpaid work.

The 1996 Census Unpaid Work Data Studies will consist of a series of three studies. This first paper is an evaluation study aimed at assessing key aspects of the *quality* of census unpaid work data. Aspects of fathering and elder care will be the subjects of two other analytical studies that Status of Women Canada expects to publish, also based on the 1996 Census data on unpaid work. The work on fathering will include additional information about patterns of sharing of child-care work by household members, other than the father and mother, thanks to the fact that the 1996 Census broke important ground in obtaining the relevant data from all members (aged 15 or more) in the sample households.

The data on unpaid work can be used to produce research results that are applicable in a wide variety of policy fields including:

- the adequacy of child-care arrangements available to meet existing needs;
- planning and delivery of assistance for those who devote large portions of their time to providing personal services to the needy elderly;
- the linkages among hospital services, community care services and family-level informal care;
- using the tax system to recognize and compensate unpaid work that is vital to the well-being of Canadians and the integrity of Canadian society;
- family-friendly workplace policies, training policies and employee benefit policies; and
- impacts on child care when the household is the major venue for the family's market production (e.g., farm households, households where the home is also the place of business for the self-employed).

A better understanding of the links between paid and unpaid work—market and non-market work—has long been identified by feminist activists and academics as critical to understanding inequality between women and men, and to developing more equitable policies. In Canada, where measurement and valuation capacity is among the best in the world, attention has shifted to policy implications. The presence of the questions on unpaid

work in the 1996 Census and the subsequent collection and analysis of the data will make a significant and important contribution to these discussions by:

- drawing attention to the existence and importance of work done outside the paid work force;
- allowing for the systematic collection of high-quality information across the country (This is especially significant for provincial and sub-provincial jurisdictions, which often can receive no data based on national surveys due to problems with small sample sizes or due to being totally omitted from the national surveys, for example, the Territories.); and
- contributing to the creation of the largest data set dealing with unpaid work in Canada as well as a richer and more useful data set as subsequent iterations of the Census of Canada capture the same information—allowing for better and more reliable analyses over time.

Status of Women Canada's objective is to enhance public debate on gender equality issues and to enable individuals, organizations, policy makers and policy analysts to participate more effectively in the development of policy. We are pleased to support this work, and we thank the analysts at Statistics Canada for their significant and timely contributions.

ACKNOWLEDGMENTS

The authors are indebted to Michel Coté and his staff for their generous support in the conduct of special retrievals of tabulations from the 1996 Census master files, and to Chris Jackson for help in correcting defects in previous draft texts. Special thanks are also due to Ian Macredie and Benoit Laroche for permitting substantial staff resources to be released for use in this study. Several of the final charts and tables are the result of the skilled handiwork of Sharron Smith.



IMPORTANCE OF THE CENSUS UNPAID WORK DATA AND MAIN ISSUES ADDRESSED IN THE EVALUATION STUDY

Introduction

No existing body of statistics is without important limitations. New kinds of statistics are especially subject to improvements as experience with their uses grows. The new 1996 Census data on aspects of unpaid work are no exception to this rule. Despite the limitations, with ingenious and imaginative applications of statistical estimation procedures (yet no more imaginative than what we now see in the commonly accepted usage of simulations and synthetic data in support of policy analysis and development), the 1996 Census data on unpaid work can be used to produce research results that are applicable in a wide variety of policy fields.

Precisely because these are brand new data in the context of a census, an essential foundation for their substantial analysis is the conduct of evaluation studies aimed at assessing key aspects of the quality of these data. First, some useful notion of "overall quality" needs to be developed and then applied in an evaluation of the census unpaid work data relative to the range of quality characteristic of census data concerning questions considered to be legitimate and established for use in a national census. If the overall quality is judged to be within that range, then it is valuable to identify major population subgroups where the quality is especially weak or better than average in some useful sense. It is also important to identify large classes of analytical applications that are blocked because of peculiarities in the census questions or in the subsequent representation of the census results on the census master file at Statistics Canada. The challenge of conducting these evaluations has been partly addressed in three recently completed studies, which are now brought together as the subsequent chapters of this document. Related information, focussed on opportunities and limitations for analysis purposes, can be found in Stone and Silver (1998a).

Once there is a positive decision to harness the census vehicle for delivering data about unpaid work (as happened with the 1996 Census), and the data are gathered, a new evaluation process of potentially national significance begins to unfold.

Main Issues Addressed

A summary of the issues addressed in each paper follows.

Paper 1 assesses the quality of the estimates for the three kinds of unpaid work covered in the 1996 Census through an examination of the non-response rates and the scale of imputations due to manifestly inappropriate responses. The question addressed is whether the census estimates of the volumes of unpaid work are of a minimum acceptable quality that would warrant their use in further analyses of unpaid work for various sub-populations of Canadian society, where quality is evaluated only from the perspectives just cited.

Paper 2 evaluates the quality of the results of the 1996 Census questions on unpaid housework, unpaid child care and unpaid care to seniors, by comparisons with data yielded by identical questions asked on the 1996 General Social Survey (GSS). The key question addressed in this evaluation is: What are the quality implications of the differences between the 1996 GSS and the 1996 Census, given that identical questions were used, when we compute similar distributions (e.g., the distribution of respondents over categories of time spent in child-care work) using the two data sources?

Paper 3 evaluates the quality of the results of the 1996 Census questions on unpaid housework and unpaid child care, by means of comparisons with data yielded by the time-diary data from the 1992 GSS. This study tests whether certain patterns of difference between diary and stylized data (the census questions fall into the latter class), as described by the literature, are found in reasonably comparable data for selected distributions from the 1992 GSS and the 1996 Census. Only the census questions on housework and child-care activity are addressed in this paper, as the GSS diary data are not suitable for an analysis of time spent on care to seniors.

Selected Findings

The following capsule summary of key findings from all the papers may be useful for the reader who does not have the time to plough through all three studies.

The first two evaluation papers have found the census unpaid work data to be within the "quality range" relative to much more established census variables. The worst case is child care—it is close to the low end of the quality range when the comparison is with the same question asked in the 1996 GSS (see Paper 2). (This "same question" is not the time-diary, child-care data discussed above.)

However, for a major segment of the population—adults aged 25 to 64 with post-secondary education, particularly those in the labour market—the census and GSS data compare very well, once the sources of non-comparability between the two data sets are considered. For this population, it is hard to say which data source is better. As there are some subtle concepts in the framing of census-type, time-use questions in connection with caring work, it is conceivable that better education means a better chance to appreciate and respond appropriately to these questions.

For certain key groups—the unemployed, those with low education, exotic ethnic/language groups—the census data should be used with great caution. However, for these same groups, the GSS sample size is often so small that neither source looks good from a quality perspective.

The response categories used in the unpaid work questions are a major weakness of the census data. A key task will be to fix the categories to make them more useful for analysis purposes.

Finally, one other advantage of the census cannot be overlooked: its large sample size and its subsequent ability to provide data at the sub-provincial level or for any number of special

sub-populations of interest. This is of particular significance as the focus of much social policy is below the national level, at the provincial, sub-provincial and even community levels.

Major Observations on the Findings

The findings developed in papers 1 to 3 give rise to a number of important general observations of conclusions that deserve to be highlighted.

First, because they deal only with housework and with caring work on behalf of children and seniors, no estimate of any defined *total volume of unpaid work* can be made from the census. Hence, debates over whether group A is doing more unpaid work than group B, after considering their sizes and other factors that must be controlled, cannot be supported by the census data. Rather, analyses about the specific sub-fields of unpaid work are what should be considered with these data. Fortunately, the 1996 Census questions include fields where there is major public interest.

Second, because census respondents provided their information in terms of broad categories of time spent doing specific classes of unpaid work, and with a substantial open-ended category included, estimates of time spent doing particular kinds of work cannot be routinely retrieved in census tabulations. Instead, statistical models have to be developed and applied to transform responses initially given in terms of broad categories into estimated numbers that can be added to produce simulated totals. With regard to the 1996 Census data on time spent doing child-caring work, such a model has already been built, tested and used (see Stone and Silver, 1998b).¹

Third, much more extensive use of the data on time spent helping seniors would be possible if there was an additional question designed to allow analysts to identify the respondents who had little or no opportunity to provide help to any senior. The lack of this additional question limits the analysis of these data to situations where a census respondent lives with a senior, especially as a family member or other relative. Fortunately, the census sample of such households is much greater and more useful to provincial jurisdictions (the main venues for policies and programs, and services to the elderly) than the sample available from any Statistics Canada survey.

Because the census also has data on the disability statuses of all the household members, this data resource becomes a centrepiece of the informational support to debates (and related program planning) related to the volume and distribution of *informal* home care activities among a wide variety of Canadian sub-populations.

Fourth, with regard to child-caring work, it is almost totally misleading to suggest that the GSS time-diary estimates of time spent doing child caring as a primary activity are comparable (let alone superior) to the census-type, stylized question ("how many hours did you spend..."). Paper 3 conclusively supports the proposition that fundamentally different variables are involved. (An analogy would be to confuse total income defined in terms of what one receives in return for labour output with total income in terms of both labour

output and many other kinds of income source such as investment income. They are simply two fundamentally different kinds of totals.)

The superiority of the census-type question arises from the fact that it implies reference to a much broader concept of child-caring work than does the GSS time-diary data that focusses on measuring primary activities only. Among child-care analysts and mothers, this broader concept is far more likely to be judged useful, than the primary-activity concept of the time-diary data. Hence, if the GSS is a serious alternative to the census, as a source of data on time spent doing child-care work, it will be through the use of the census-type question in a GSS (and not via the GSS time-diary data, unless these are accompanied by a supplementary child-care diary). This is an important conclusion, because the major opposition to unpaid work data in the census, on grounds of defective quality, is based on the alleged superiority of time-diary data. We are saying that, on the contrary, as far as child care is concerned, time-diary data that measure only primary activities are an inferior substitute for the census-type question.

It follows that for access to time-use data involving the more comprehensive and useful concept of child-caring work in a context where a variety of key population subgroups need representation through adequate sub-sample sizes, the national surveys offer no substitute for the census. The census is the only source that can deliver the needed information.

Finally, taking note of the fact that the GSS is the vehicle thought to provide data on time use that are superior to what can be obtained in a census, the papers show that there is a very large sub-population of Canadians where the alleged superiority of the GSS data is extremely difficult to establish with regards to the three items covered in the census (to state the case most favourably for the GSS).

PAPER 1: NON-RESPONSE AND IMPUTATION FOR MULTIPLE RESPONSES

Introduction

This paper assesses, in general terms, the quality of the estimates for the three kinds of unpaid work covered in the 1996 Census through an examination of the non-response rates and imputation of the data. The question addressed is whether the census estimates of the volumes of unpaid work are of a minimum acceptable quality to warrant their use in further analyses of unpaid work for various sub-populations of Canadian society. (For a summary of the findings, see the end of this paper.)

Background

Questions on unpaid work were included on the census for the first time in 1996. Information was collected for a sample of one in five households using the 2B (long form) questionnaire.² All respondents aged 15 and over, excluding full-time residents of institutions, were asked the questions on unpaid work.

There were three questions. Respondents were asked to report the number of hours spent in the week prior to enumeration, doing (1) *unpaid housework*, (2) *unpaid child care* and (3) *unpaid care or assistance to seniors*. Respondents were to check the box for the category which contained the number of hours they spent doing each activity. (See Appendix 1.A for a copy of the questions as they appeared on the questionnaire and the accompanying instructions from the guide.)

An important concept here is that of "*overlapping activity*." This phrase points to periods of time when the respondent was simultaneously carrying on activities in at least two of the three areas named above. For example, doing housework and child care at the same time would represent overlapping activity. In instances of overlapping activities, respondents were asked to report the time spent in both areas, e.g., counting the same time in both the housework and the child-care questions.

Data Quality

Once the completed questionnaires were returned from the field, the responses were data captured at one of the seven Revenue Canada processing centres across Canada. The captured data was then transferred to Statistics Canada to be loaded onto a data base for further processing.

Before processing begins, the incoming data are examined for indications of respondent comprehension and difficulties answering the questions, as well as any resistance to potentially sensitive or intrusive questions. Two indicators used to evaluate these problems are the non-response rate and the multiple response rate where only one response is appropriate.

Non-Response Rate

The non-response rate is defined as the number of persons in a given category of the population who did not answer a particular census question *when they should have*, divided by the total population of that category who should have answered the question. For example, since all respondents aged 15 and over (excluding institutional residents) should have answered the questions on unpaid work, the non-response rate for a given unpaid work question (e.g., housework) would be calculated as:

The number of persons 15 and over with no response for unpaid housework

The total number of persons aged 15 and over

Non-response rates can be calculated for each question appearing on the census questionnaire. They can also be calculated for any subgroup of interest in the population, for example, all residents of Ontario or persons aged 65 and over.

A high non-response rate (i.e., one that is generally higher than that of other census questions or population subgroups) can indicate one or more potential problems. Respondents might not have understood the question clearly or found the question too difficult and, therefore, were unable to provide an answer. Or, it can indicate that respondents found the question so intrusive or sensitive that they refused to answer (e.g., common-law status or income).

Finally, respondents may not answer a question because they think it does not apply to them.

For example, the marital status question is asked of all household members, even young children. However, many respondents leave this question blank for their children (marital status is not applicable to children) rather than indicate SINGLE.

Table 1.1 presents the unweighted³ non-response rates for the three unpaid work variables in comparison with the non-response rates for several other census questions. The selected questions vary in their degree of difficulty and sensitivity for the respondents.

Looking at the rates for the unpaid work questions, the data show the question on unpaid housework to have the lowest non-response rate at 1.5%, while the question on unpaid child care has the highest rate at 3.9%. Non-response for the question on care of seniors is 2.6%.

Compared to the other census questions, the one on unpaid housework has a relatively low rate of non-response, while the question on unpaid care to seniors falls more or less in the middle. The rate for unpaid child care, by contrast, is relatively higher, approaching that seen for the questions on common-law status and ethnic origin, although not as high as the rates for industry or income—questions which many respondents find difficult to answer.

Table 1.1 Comparison of the Non-Response Rates of the Unpaid Work Questions with Selected Census Questions (20% sample data – unweighted)

Question no.	Question	Non-response rate
Q5	Marital status	1.9*
Q6	Common-law status	4.2*
Q13	Place of birth	0.9
Q17	Ethnic origin	4.1
Q24	Elementary/secondary school	2.9
Q25	Years of university	1.7
Q26	Years of college	2.2
Q30A	Unpaid housework	1.5
Q30B	Unpaid child care	3.9
Q30C	Unpaid care to seniors	2.6
Q31	Hours of paid work	1.5
Q34	Looked for work	2.9
Q36	When last worked	2.7
Q37/38	Industry	4.6
Q45	Weeks worked in 1995	3.9
Q47	Income	5.2

Note:

* Based on 100% data from the 1996 Census.

The low non-response rate for the housework question suggests respondents generally did not have difficulty understanding or answering this question, nor did they find it too intrusive. Given this result, it is unlikely that respondents would have judged the questions on child care or care to seniors as being unduly sensitive. Although many respondents may have had trouble deciding which particular behaviours constituted caring, our hypothesis is that the higher non-response rates for these two questions, in particular the one on unpaid child care, is more likely attributable to the tendency for some respondents to skip questions they think do not apply to them. For example, persons without children may have left the child-care question blank rather than check the NONE box, although some of these persons may have provided unpaid child care to the children of others.

Table 1.2 presents the unweighted non-response rates for the unpaid work variables, comparing the rates of men and women in each province and territory. It shows that:

- Men had higher non-response rates than women.
- Yukon had the highest rates of non-response for all three variables.
- The Northwest Territories had the second highest rate for the housework question; however, British Columbia had the second highest rate for the questions on unpaid child care and care of seniors.
- Newfoundland, Quebec and New Brunswick generally had the lowest rates of non-response for the three variables.

To elaborate on these points, the fact that men had consistently higher non-response rates than women may be further evidence of our hypothesis that some respondents skip questions they think don't apply to them. It would not be surprising to find some respondents who thought the questions on unpaid work applied only to women.

**Table 1.2 Non-Response Rates, by Sex for Unpaid Work Variables
for Canada, Provinces and Territories
(20% sample data – unweighted)**

	Housework			Child care			Care of seniors		
	Total	Male	Female	Total	Male	Female	Total	Male	Female
Canada	1.5	1.6	1.4	3.9	4.1	3.6	2.6	2.7	2.4
Newfoundland	1.0	1.0	1.0	2.9	3.0	2.8	1.7	1.7	1.6
Prince Edward Island	1.4	1.5	1.2	4.2	4.5	3.9	2.6	2.8	2.4
Nova Scotia	1.3	1.3	1.2	4.2	4.5	3.9	2.4	2.5	2.4
New Brunswick	1.4	1.5	1.3	3.6	4.1	3.2	2.2	2.4	2.0
Quebec	1.3	1.3	1.2	2.7	2.8	2.5	2.2	2.2	2.2
Ontario	1.5	1.7	1.4	4.3	4.6	4.0	2.7	2.9	2.5
Manitoba	1.2	1.3	1.1	3.4	3.7	3.2	2.0	2.2	1.9
Saskatchewan	1.2	1.3	1.2	3.6	3.8	3.4	2.2	2.3	2.1
Alberta	1.6	1.7	1.4	4.2	4.5	3.8	2.5	2.7	2.3
British Columbia	1.9	2.1	1.8	5.0	5.4	4.7	3.3	3.6	3.1
Yukon	5.1	5.6	4.4	6.7	7.4	5.9	5.7	6.4	4.8
Northwest Territories	2.2	2.3	2.1	3.2	3.4	3.1	2.7	2.8	2.6

**Table 1.3 Comparison of the Non-Response Rates of the Unpaid Work Questions with
Selected Census Questions, Canada, Provinces and Territories
(20% sample data – unweighted)**

	Place of birth	Years of university	Years of college	Unpaid house-work	Unpaid child care	Unpaid care of seniors	Hours of paid work	Looked for work	When last worked	Industry	Weeks worked in 1993
Canada	0.9	1.7	2.2	1.5	3.9	2.6	1.5	2.9	2.7	4.6	3.9
Newfoundland	0.3	0.9	1.3	1.0	2.9	1.7	1.1	1.8	1.4	4.4	3.4
Prince Edward Island	0.6	1.6	2.2	1.4	4.2	2.6	1.6	3.0	2.6	4.4	3.6
Nova Scotia	0.5	1.4	1.9	1.3	4.2	2.4	1.3	2.4	2.0	4.6	3.5
New Brunswick	0.9	1.5	1.9	1.4	3.6	2.2	1.4	2.6	2.2	4.7	3.6
Quebec	0.7	1.4	2.0	1.3	2.7	2.2	1.3	2.7	2.4	3.9	3.7
Ontario	1.0	1.8	2.3	1.5	4.3	2.7	1.6	3.1	2.8	5.1	4.0
Manitoba	0.6	1.4	1.8	1.2	3.4	2.0	1.3	2.6	2.2	3.7	3.2
Saskatchewan	0.6	1.4	1.8	1.2	3.6	2.2	1.4	2.9	2.5	4.4	3.4
Alberta	0.8	1.7	2.1	1.6	4.2	2.5	1.6	3.0	2.7	4.4	3.8
British Columbia	1.2	2.2	2.7	1.9	5.0	3.3	2.0	3.6	3.3	5.2	4.7
Yukon	3.7	5.2	5.6	5.1	6.7	5.7	4.8	7.0	6.1	6.7	6.4
Northwest Territories	1.5	2.6	3.1	2.2	3.2	2.7	2.2	3.9	3.5	3.2	3.5

However, the possibility that men do not identify their household work with the questions on unpaid work has implications for the quality of the census estimates: Is the amount of time spent by men on unpaid work underreported? The distribution of the census data for men will be of particular interest when we examine the quality of the census estimates in subsequent reports.

The second observation regarding the high non-response rate for Yukon is not a situation that is unique to the questions on unpaid work. As Table 1.3 shows, in all the census questions for which provincial data were available, the non-response rates for Yukon were the highest. Similarly, the next highest rates alternated between the Northwest Territories and British Columbia. Thus, while we should be cautious about using unpaid work data from the Yukon, this caution applies to other census variables as well.

Multiple Response Rate

The second indicator used to evaluate respondents' comprehension and reaction to a particular census question is the multiple response rate for an item where only a single response is appropriate. Erroneous multiple responses occur when a respondent provides more than one valid answer to a question but only one response is appropriate. (In the case of unpaid work, the respondent checked more than one "hours" category for a given unpaid work question.) This concept gives rise to the definition of an "erroneous multiple response rate," which is the number of erroneous multiple responses per 10,000 responses made.

Table 1.4 Comparison of the Multiple Response Rates of the Unpaid Work Questions with Selected Census Questions (20% sample data – unweighted)

Question no.	Question	Multiple response rate (per 10,000 responses)
Q5	Marital status	10*
Q13	Place of birth	28
Q30A	Unpaid housework	77
Q30B	Unpaid child care	39
Q30C	Unpaid care to seniors	18
Q34	Looked for work	31
Q36	When last worked	63
Q41	Class of worker	51

Note:

* Based on 100% data from the 1996 Census.

Table 1.4 shows the unweighted multiple response rate for the census unpaid work questions in comparison with other census questions similar in format to those on unpaid work (i.e., questions where the respondent was to check one box from a selection).

A key observation is that the absolute level of erroneous multiple responses is, in fact, extremely low—scarcely high enough to affect any statistic based on a sample of at least 1,000 respondents. For example, the question on unpaid housework had the highest rate at about 77 erroneous multiple responses per 10,000

responses, while the questions on child care and care of seniors had approximately 39 cases per 10,000 and 18 cases per 10,000 respectively.

Although the absolute levels of erroneous multiple response rates are very low, the rate for responses concerning unpaid housework are the highest among the set of census variables for which data are available. (The next highest rate was 63 cases per 10,000 responses for the question on when last worked.) In contrast, the rates for the other two unpaid work questions are close to the lowest figures shown in Table 1.4. These figures pertain to marital status and place of birth, where one would expect extremely low levels of erroneous multiple responses. These observations indicate data of acceptable quality, from the perspective of this particular indicator, in at least two of the census unpaid work questions.

There is no implication here that the data on unpaid housework should be regarded as being of bad quality, since the absolute level of the indicator is extremely low. However, these results may be a warning that this variable deserves special attention as we develop our upcoming appraisals of data quality using other indicators.

The higher multiple response rate for housework could indicate some difficulty on the part of respondents in calculating the number of hours spent doing this activity. This question encompasses a more varied range of activities than either the child-care or care-to-seniors question, and respondents may be unsure which activities to include or have difficulty recalling the time spent on them. Also, if respondents had first included the time spent on child-care or senior-care activities in their response to the housework question, they may have had to go back and correct their first response. If this correction was not observed by the person capturing the data from the questionnaire, a multiple response would have been captured.

Edit and Imputation

The process by which non-response and multiple response errors are corrected in the census is termed "edit and imputation."

Editing involves inspecting the captured responses in search of errors such as non-response and erroneous multiple response. As already described, these errors can be the result of respondents answering the questions incorrectly or incompletely, or they can be due to data entry errors generated during the data capture from the questionnaires.

Imputation is the process of inserting alternate or new responses designed to correct the errors detected during editing. There are two broad classes of imputation: deterministic imputation and stochastic imputation. In deterministic imputation, errors are corrected by inferring the appropriate value from answers to other census questions. For example, if there is no response to the question about sex, but the person is described elsewhere as being the son of Person 1 then "male" can be imputed as the correct response for sex.

To explain stochastic imputation, consider the example of an item that is missing or clearly incorrect, which cannot be imputed using deterministic imputation. Suppose a person's captured responses indicate that he/she is employed at the Ottawa Hospital, but the question on occupation has no entry. In principle, stochastic imputation involves randomly searching the already corrected census records for another person who is regarded as being "sufficiently similar in relevant respects" and who does have an entry for occupation. That person's response is then entered into the record where occupation was missing. The person supplying that information is called a "donor." The phrase "donor imputation" is used below to refer to this kind of imputation.

The phrase "sufficiently similar in relevant respects" needs clarification. Again, for simplicity, let us proceed with our example. Since we already know that the person is working at the Ottawa Hospital, we might decide that working at a hospital is a relevant respect and, therefore, only persons so employed would be eligible to become donors. In general, a set of

characteristics or "matching criteria" thought to be related to the missing item is selected, and a search is made for persons who match the record in error on those characteristics. These persons are considered to be donors. The missing information is then imputed from among the donor records.

From time to time, a complete match on all the criteria is not found. In that situation, the best possible match is found within the established limits of the search.⁴

The foregoing remarks were designed to introduce the reader to the main features of imputation. In the next few paragraphs, we describe how imputation was applied to the kinds of errors discussed above.

Imputation of Non-Response

Respondents 15 years and over, who were missing a response for one or more unpaid work questions, were assigned a value through stochastic imputation. The imputation of the unpaid work variables was divided into two parts. Questions on unpaid housework and child care were imputed in a separate process from that dealing with care of seniors. It was felt that, in selecting a donor, characteristics that might determine the number of hours spent on housework or child care, e.g., the presence of children in the household, would be less relevant to a question on the number of hours spent caring for seniors. Separating the imputation process into two parts allowed for the use of different matching criteria for the two sets of variables.

To find a donor to impute missing housework and child-care data, records were first stratified by type of dwelling. Persons living in collective dwellings were further stratified according to the type of collective they lived in. Persons in private dwellings were stratified by sex and age group. Within these dwelling type and sex and age strata, the following variables were used as additional matching criteria in the choice of donor: five-year age groups, marital status, presence of children, number of children, number of hours of paid work and place of work.

The imputation process for the care of seniors variable stratified persons living in collective dwellings according to the type of collective. Persons in private dwellings, however, were stratified by sex and membership in an economic family with a senior member (defined as persons aged 60 and over). Persons in economic families without a senior member were further divided by age into persons under 45 and persons 45 and over. The matching criteria used to select a donor were five-year age groups, marital status, place of work, number of hours of paid work and age group of senior (for persons living with a senior).

As a result of the imputation processes, a donor who fully matched the erroneous record on all matching criteria was found in approximately 92% of cases⁵ in the imputation of the housework and child-care variables, and in 98% of cases in the imputation of the care of seniors variable. In both processes, donors could be found for all the records.

Table 1.5 shows the distribution of the unpaid work variables before and after stochastic imputation has taken place. The table shows minimal change in the distributions as a result of the imputation procedures used. These data are for Canada as a whole. It is worthwhile to note that, for these sub-populations, there is minimal change in the distributions resulting from the imputation process.

Resolution of Erroneous Multiple Responses

Multiple response errors for the unpaid work questions were resolved through deterministic imputation. In all cases, the category with the lowest hours marked by the respondent was assigned to the respondent. The lowest category was chosen to avoid assigning a disproportionate number of persons to the highest "hours" category, while still retaining one of the original responses from the questionnaire.

Table 1.6 shows the distributions of the data before and after the deterministic imputation took place. As the number of multiple response errors was extremely small, compared to the total number of responses, their imputation has had a minimal effect on the distribution of the data.

Conclusion

This report begins the evaluation of the quality of the estimates of unpaid work collected from the 1996 Census through an examination of the non-response and multiple response rates and subsequent imputation of the data. The results of the examination found that, compared with other census questions, the non-response rate for the question on unpaid housework was relatively low. However, the rates for the questions on unpaid care to seniors, in particular, unpaid child care, were comparatively higher. It is our hypothesis that the higher non-response

Table 1.5 Comparison of the Unpaid Work Variables Before and After Imputation of Non-Response, Canada
(20% sample data – unweighted)

	Before imputation	After imputation of non-response	Ratio of distribution before and after imputation
Unpaid housework:			
Total	100.0	100.0	
None	11.9	12.0	1.01
Less than 5 hours	22.2	22.3	1.00
5-14 hours	29.9	29.9	1.00
15-29 hours	19.2	19.1	0.99
30-59 hours	11.6	11.6	1.00
60 hours or more	5.2	5.2	1.00
Unpaid child care:			
Total	100.0	100.0	
None	60.4	60.9	1.01
Less than 5 hours	9.7	9.6	0.99
5-14 hours	10.3	10.2	0.99
15-29 hours	7.0	6.9	0.99
30-59 hours	5.4	5.4	1.00
60 hours or more	7.2	7.1	0.99
Unpaid care of seniors:			
Total	100.0	100.0	
None	83.0	83.0	1.00
Less than 5 hours	10.9	10.8	0.99
5-9 hours	3.4	3.4	1.00
10 hours or more	2.7	2.7	1.00

rates for the latter questions are more likely attributable to the tendency for some respondents to skip questions they think do not apply to them.

Men had consistently higher non-response rates than women for all the provinces and territories. While this may just be further evidence of the hypothesis stated above, the question is raised as to whether men are less likely to identify their household work with the census questions on unpaid work and, ultimately, underreport the amount of time they spend on these activities.

Evaluation of the erroneous multiple response rates indicates data of acceptable quality for the child-care and care-to-seniors variables. In contrast, the rate for responses concerning unpaid housework was the highest among the set of census variables for which data were available. However, since the absolute level of multiple response is extremely low, the data on unpaid housework should not be regarded as being data of bad quality. Rather, these results may be a warning that this variable deserves special attention in future analyses.

Table 1.6. Comparison of the Unpaid Work Variables Before and After Imputation of Multiple Response
(20% sample data – unweighted)

	Before imputation	After imputation of non-response	Ratio of distribution before and after imputation
Unpaid housework:			
Total	100.0	100.0	
None	11.7	11.9	1.02
Less than 5 hours	22.2	22.2	1.00
5-14 hours	30.0	29.9	1.00
15-29 hours	19.2	19.2	1.00
30-59 hours	11.6	11.6	1.00
60 hours or more	5.3	5.2	0.98
Unpaid child care:			
Total	100.0	100.0	
None	60.5	60.4	1.00
Less than 5 hours	9.7	9.7	1.00
5-14 hours	10.3	10.3	1.00
15-29 hours	6.9	7.0	1.01
30-59 hours	5.4	5.4	1.00
60 hours or more	7.2	7.2	1.00
Unpaid care of seniors:			
Total	100.0	100.0	
None	83.0	83.0	1.00
Less than 5 hours	10.9	10.9	1.00
5-9 hours	3.4	3.4	1.00
10 hours or more	2.7	2.7	1.00

Deterministic imputation to the lowest "hours" category marked by the respondent was used to resolve cases of erroneous multiple response. Stochastic or donor imputation was used to correct non-response. The imputation process had a minimal effect on the resulting distributions of the data for both Canada and the provincial subgroups of the population.

Appendix 1.A: Census Unpaid Work Questions and Guide Instructions

<u>HOUSEHOLD ACTIVITIES</u>	43.
<p>Note: Last week refers to the seven days (Sunday to Saturday) before Census Day.</p> <p>In Question 30, where activities overlap, report the same hours in more than one part.</p> <p>30. Last week, how many hours did this person spend doing the following activities?</p> <p>(a) Doing unpaid housework, yard work or home maintenance for members of this household, or others.</p> <p>Some examples include: preparing meals, doing laundry, household planning, shopping and cutting the grass.</p>	<p>01 None</p> <p>02 — Less than 5 hours</p> <p>03 — 5 to 14 hours</p> <p>04 — 15 to 29 hours</p> <p>05 — 30 to 59 hours</p> <p>06 — 60 hours or more</p>
<p>(b) Looking after one or more of this person's own children, or the children of others, without pay.</p> <p>Some examples include: bathing or playing with young children, driving children to sports activities or helping them with homework, and talking with teens about their problems.</p>	<p>07 — None</p> <p>08 — Less than 5 hours</p> <p>09 — 5 to 14 hours</p> <p>10 — 15 to 29 hours</p> <p>11 — 30 to 59 hours</p> <p>12 — 60 hours or more</p>
<p>(c) Providing unpaid care or assistance to one or more seniors.</p> <p>Some examples include: providing personal care to a senior family member, visiting seniors, talking with them on the telephone, and helping them with shopping, banking or with taking medication.</p>	<p>13 — None</p> <p>14 — Less than 5 hours</p> <p>15 — 5 to 9 hours</p> <p>16 — 10 hours or more</p>

Guide Instructions**HOUSEHOLD ACTIVITIES**

Question 30 will provide information on how much time people 15 years and older spend at household tasks, on caring for children and in providing care and assistance to elderly persons. This information will provide a better understanding of how these unpaid activities contribute to the well-being of Canadians.

QUESTION 30 – Unpaid Activities

Mark the circle that contains the total number of unpaid hours spent last week doing each of the activities in parts (a) to (c).

Include hours spent doing unpaid activities for:

- members of one's own household;
- other family members outside the household;
- friends or neighbours.

Do not include hours spent:

- working for pay (report paid work in Question 31);
- doing unpaid volunteer work for a non-profit or religious organization, charity or community group.

Overlapping Activities

People often perform more than one unpaid activity at the same time. For example, a person may spend one hour preparing a meal while at the same time looking after his/her children. This person should report one hour of housework in part (a) and one hour of child care in part (b) of Question 30. It does not matter that these activities took place at the same time.

Part (a) – Doing unpaid housework, yard work or home maintenance

No further instructions.

Part (b) – Looking after children without pay

Report hours spent doing activities such as talking or playing with children if, during these activities, this person was responsible for their care.

Part (c) – Providing unpaid care to seniors

Seniors are all persons 65 years of age and over and some individuals close to 65 suffering from age-related infirmities.

PAPER 2: QUALITY OF THE 1996 CENSUS UNPAID WORK DATA IN LIGHT OF COMPARISONS WITH 1996 GSS DATA BASED ON IDENTICAL QUESTIONS

Introduction

This paper evaluates the quality of the results of the 1996 Census questions on unpaid housework, unpaid child care and unpaid care to seniors, by means of comparing data yielded by identical questions asked on the 1996 General Social Survey (GSS). The key question addressed in this evaluation is: What are the quality implications of the differences between the 1996 GSS and the 1996 Census, given that identical questions were used, when we compute similar distributions (e.g., the distribution of respondents over categories of time spent in child-care work) using the two data sources?

This paper is the second in a series of reports on the evaluation of the quality and "usability" of the census questions on unpaid work. The first paper looked at quality of the census data in terms of non-response rates and levels of imputation caused by manifestly inappropriate responses (Swain and Stone, 1998). The next report presents results of a quality assessment based on using the data from the 1992 GSS time-use diaries.

Methodology

The Dissimilarity Index (DI) is the key statistic used to conduct the comparisons presented below. It involves computing a percentage distribution (on a given variable) from the GSS data, repeating the computation for the same variable from the census data, and then gauging the overall discrepancy between the two distributions. The DI is one half the sum of the absolute differences between the apparently comparable distributions drawn from the two surveys. This is a well-known index among statisticians.

The discussion begins by considering the level of the DI for the entire census and GSS samples (representing the population aged 15 or more). With this value in hand, the following text then considers whether it indicates serious quality defects in one or both of the two data sources.

It is possible to raise and answer some standard statistical questions when trying to decide whether a given level of the DI implies "significant" divergence between the two data sources. A test of statistical significance, based at least on bootstrapping techniques that hold the GSS distribution constant, is possible. This would allow one to answer the question as to whether the DI level could easily have arisen by chance, if the GSS distribution is the correct one. We avoid constructing such a test because, in most cases, the census sample size, at least, is so large that non-substantial DI values could appear to be significant. Even though only statistical significance would be at issue, many users of these results would subtly extend the importance of the results to the realm of substantive significance.

Therefore, we use two other approaches to create a perspective from which to assess whether a given DI value should be taken as indicating a serious discrepancy between the census and the GSS. The first approach involves consideration of the following question. Suppose we take the GSS distribution as being correct. At what level of misclassification of responses in the census data would we begin to feel that we are in great danger of being misled by the census data?

The answer to this question is arbitrary (as is the selection of a level of significance in a formal test of significance). However, we note that in the official data release guidelines a coefficient of variation of 15% or less is thought to identify data of sampling variability low enough to be published as reasonably reliable. Using this practice as a cue, only when the DI exceeds 15% should we suggest that one of the two data sources is probably seriously damaged by response errors.

Of course, a DI value of 15% should be considered large in some useful sense. However, as we argue below, it is wrong to conclude quickly that the problem lies entirely on the side of the census data.

The second approach to developing a perspective for interpreting a single value of the DI involves computing DI levels for more common demographic variables (e.g., age or marital status). By selecting variables known to be measured fairly accurately, we can create one basis for deciding whether a given level of DI in the data on unpaid work probably points to serious damage to the quality of data in either of the two sources.

The study of variations among selected population subgroups is a major activity supported by census data. Thus, the paper presents results of computing DI values for several subpopulations. However, the key purpose of looking at these subgroup-level variations is to help indicate populations where the census data seem especially weak and others where they seem to stand up well in comparison with the GSS data. Thus, in the discussion of the subgroup variations, we selectively pinpoint particular populations that stand out in terms of apparent weakness or strength in quality of the census data. There is no effort to review the pattern of variation in DI values over various breakdowns of the overall population. The discussion, therefore, is very limited when compared to the scope of the information provided in the tables of DI values shown below.

The presentation is in three major parts, one for each of the census unpaid work questions. Before showing any data, however, it is necessary to raise some important cautions regarding the alleged superiority of the GSS data. These cautions arise from the major sources of non-comparability in methodology and time reference between the 1996 Census and the 1996 GSS.

Comparability of the Census and the GSS

The 1996 GSS is the only survey to contain the three unpaid work questions as they appeared on the 1996 Census. Thus, it provides a clear opportunity to compare and evaluate the

quality of the census results. Of the two surveys, the GSS is generally believed to be the superior data source for several reasons outlined below.

Users of the data should be cautious in interpreting differences between the GSS and the census as indications of the inferiority of the census data. Although the census and the GSS used identical questions to collect information on unpaid work, there are several differences between the two surveys that can have a direct impact on the results obtained such that the comparability of the two sources can be called into question.

The first major difference is the method of data collection. The GSS is conducted by an interviewer over the telephone, while the census is conducted through self-enumeration. It is well known that this difference alone can affect data where a respondent might change her/his response to a telephoned question if given time to think for several minutes about the correct answer. For example, when asked "what was your income in the last calendar year," you might give one response if you have 15 seconds to reply and an entirely different one if you have time to consult your tax records.

The GSS responses are believed to be more reliable, because the questionnaire is administered under the control of a trained interviewer who is immediately available to help respondents with any difficulties they have in understanding or answering the unpaid work questions. The interviewer can advise respondents as to what activities should be included under each of the unpaid work categories and help them calculate the time spent on these activities.

On the other hand, although census respondents answer the census questionnaire without the assistance of an interviewer, self-enumeration allows them to take time to consider their answers and to consult with other household members. Without an interviewer waiting on the telephone, there is less pressure to come up with an answer quickly. For respondents who have trouble answering the unpaid work questions, a guide with additional information is provided. (It is widely believed that the guide is rarely consulted.) There is also a telephone assistance system in place where a respondent can speak to a census representative directly and receive the same kind of help provided by an interviewer. However, the respondent might not seek this help as often as he/she would if there was an interviewer involved.

A second major difference between the census and the GSS is the reference period for which the unpaid activities are reported. The GSS collects data over a period of a year, whereas the census collects data at one point in time. The GSS was conducted monthly from February to December 1996. Respondents reported their unpaid work activities for the week preceding the survey interview. The census questions on unpaid work reflect only those activities that took place in the week prior to Census Day (May 14).

Thus, seasonal variations in the variety of activities represented, as well as in the lengths of time people tend to spend doing certain activities, are reflected in the GSS data, but not in the census data. While this is certainly an advantage of the GSS data for representing the situation for a whole year, it is a disadvantage for this study where, ideally, data are required

that represent the situation for a narrowly defined reference period that is common for all respondents. Ultimately, the comparability of the census and the GSS estimates is compromised.

Another reason the GSS data are believed to be more accurate than census data is that the GSS did not allow proxy reporting for the three unpaid work questions. The GSS collects data on only one respondent from each household. The selected respondent reported time spent on each of the unpaid work activities for her/himself only. If the selected respondent was not available to complete the interview, the unpaid work questions were not asked. In the census, on the other hand, it is assumed that one household member answers the unpaid work questions—and the other census questions—for all members of the household (proxy reporting).

The comparative impacts of proxy reporting cannot be evaluated for these data sources because the GSS made no attempt to obtain time-use information about all members of the respondents' households. Both surveys should have collected data for all members of the sample households, so the two situations could be compared. Until the GSS collects data for all members of the households sampled, no one knows what would be the comparable level of proxy reporting in the GSS.

Moreover, in the census, the respondent filling out the questionnaire has time to consult other household members as needed. Also, more than one person can be involved in filling out the census questionnaire. This would not be possible in the GSS, as long as it uses telephone interviews to gather data.

In reality, no one has an estimate of the amount of census proxy reporting in which the person making the report failed to do suitable checking with other household members before filling out the form on their behalf.

One aspect in which the GSS cannot be said to be superior to the census is in sample size. With its large sample size, census estimates will be more reliable than the GSS in terms of sampling variability for dozens of significant sub-populations. The GSS sample consisted of approximately 12,000 persons, 15 years of age and over, excluding institutional residents, living in the 10 provinces. The census sample consisted of one in five households in Canada. (The census unpaid work questions were asked of household members aged 15 and over, excluding institutional residents.) Thus, the census sample is much larger than that of the GSS, leading to dozens of sub-populations for which the census has the potential to yield estimates with low sampling variability, while the GSS would have no chance of doing so.

Finally, the treatment of non-response differs significantly between the two surveys, such that the comparability of the census-GSS estimates may be affected. The census imputes all question non-responses, while the GSS imputes only key demographic variables such as age, sex and living arrangements. As a result, almost every variable in the GSS data base has an element of non-response associated with it. For some sub-populations, the level of non-response is high enough that the accuracy of the GSS estimates could be called into question.

At the very least, the GSS non-response rate should be taken into consideration whenever large census-GSS variances are observed. Where the GSS non-response rate is a substantial percentage of the overall census-GSS difference, we have reasons to be sceptical of the assumption that the GSS data are better.

In short, we support the view that the GSS probably has more accurate data for broad aggregate national-level estimates than the census, although, just how much more accurate is a question for which no answer is available due to the differences in survey methodology cited above.

When many sub-populations are being considered, the sources of non-comparability between the two sets of data are so substantial that it is unwise to assume that the GSS has superior data. And variations among sub-populations are critically important in a very large area of social science. Whenever such variations are a major aspect of analysis, the assumption that the GSS provides superior data should be regarded with scepticism.

Moreover, it is to be doubted that the value of the census unpaid work data is primarily in the production of estimates for broad, national-level aggregates, where the advantage probably lies with the GSS. It is arguable that, given what is already known from the GSS data about those aggregates, more estimates for broad national aggregates are substantially less useful and valuable than are new data for sub-populations where the GSS has no hope of yielding any usable estimates.

Global Differences

For all three unpaid work variables, the global DI values fall below the 15% level suggesting a serious census-GSS discrepancy (Table 2.1). The greatest discrepancy is indicated for the data on child care with a DI value of 11%. When compared against the DI values computed for a "comparison set" of demographic variables (Table 2.1), only the DI value for the highest level of schooling is larger at almost 13%. The other variable with a relatively large DI value is household income (11%). At 9%, unpaid housework follows household income in terms of size of census-GSS discrepancy, while the DI value for the unpaid care-to-seniors data (2%) is among the lowest values, comparable with that of marital status.

Thus, based on these global comparisons, it seems that the census unpaid work variables were measured with a level of accuracy within the range established for other variables regularly measured in the census.

As noted above, the DI values pertaining to unpaid housework and unpaid care for seniors are 9% and 2%. Among women, the figures are 11% and 3%, respectively. These DI values are higher than the corresponding ones for men, a result that is inconsistent with the belief that the census data for men are especially weak compared to similar GSS data.

Table 2.1 Global Dissimilarity Indexes Comparing Census and GSS Distributions for a Comparison Set of Census Variables

Dissimilarity Index for Census and GSS	Census	GSS	Difference
Unpaid housework:			
None	11.5	7.1	4.4
Less than 5 hours	22.7	22.9	0.2
5 to 14 hours	30.3	37.8	7.5
15 to 29 hours	19.2	21.1	1.9
30 to 59 hours	11.4	9.4	2
60 hours or more	4.8	1.6	3.1
Not stated*		4	
Dissimilarity Index			9.5
Unpaid child care (households with at least one child less than 15):			
None	18.7	13.2	5.5
Less than 5 hours	12.7	9.4	3.3
5 to 14 hours	20.6	18.1	2.5
15 to 29 hours	16.5	20	3.5
30 to 59 hours	13.5	19.3	5.8
60 hours or more	18	20	2
Not stated*		3.3	
Dissimilarity Index			11.2
Unpaid care of seniors:			
None	83.6	81.1	2.4
Less than 5 hours	10.8	11.5	0.7
5 to 9 hours	3.2	3.7	0.4
10 hours or more	2.4	3.7	1.3
Not stated*		2.8	
Dissimilarity Index			2.4
Age groups:			
Age 15-24	17	17	0
Age 25-44	41.2	41.7	0.5
Age 45-64	27.3	26.9	0.4
Age 65+	14.5	14.5	0
Dissimilarity Index			0.5
Marital status (census estimates based on 100% data):			
Never married	27	25.6	1.4
Married/common law	59.1	61.6	2.5
Separated/divorced	7.7	6.8	0.9
Widowed	6.2	6	0.2
Not stated*		0.4	
Dissimilarity Index			2.5

Table 2.1 (Continued)

Dissimilarity Index for Census and GSS	Census	GSS	Difference
Place of birth:			
In province of residence	64.4	65.8	1.3
Outside province of residence	14.2	14.4	0.2
United States	1	1.3	0.3
Central and South America	2.3	2	0.3
United Kingdom	2.9	3.3	0.4
Other Europe	7.3	6.9	0.4
Africa	1	0.9	0.1
Asia	6.6	5.2	1.4
Oceania and other	0.2	0.2	0
Not stated*		3.3	
Dissimilarity Index			2.2
Home language:			
English	66	64.9	1.1
French	22.7	23.4	0.7
Other language	9.3	5.4	3.9
English and French	0.4	1.7	1.3
English and other	1.4	3.8	2.4
English, French and other**	0.2	0.9	0.7
Not stated*		3.4	
Dissimilarity Index			5.1
Highest level of schooling:			
University	13.3	15.2	1.9
College diploma	16.4	9.7	6.7
Trades certificate/diploma	10.5	13.7	3.2
Some post-secondary	10.8	15.8	5
High school certificate	14.3	16.8	2.5
Elementary/some high school	34.8	28.9	5.9
Not stated*		3.7	
Dissimilarity Index			12.6
Occupation: (1)			
A. Management	9.4	10.8	1.4
B. Business, finance and administration	19.4	17	2.3
C. Natural and applied sciences	5.1	5.3	0.2
D. Health	5.3	6.4	1.1
E. Social science, education, government and related	7	7	0
F. Art, culture, recreation and sport	2.7	2.9	0.2
G. Sales and service	25.8	25.5	0.3
H. Trades, transportation and equipment operators	13.5	13.4	0.1
I. Primary industry	4.5	3.7	0.8
J. Processing, manufacturing and utilities	7.4	7.9	0.6
Not stated*		3.6	
Dissimilarity Index			3.5

Table 2.1 (Continued)

Dissimilarity Index for Census and GSS	Census	GSS	Difference
Total personal income:			
Without income	7.6	8	0.4
Under \$5,000	13.9	10.1	3.9
\$ 5,000 - \$ 9,999	11.8	10.7	1
\$10,000 - \$14,999	12.9	11.4	1.5
\$15,000 - \$19,999	9.2	7.8	1.4
\$20,000 - \$29,999	15	17.1	2.2
\$30,000 - \$39,999	11.4	12.6	1.2
\$40,000 - \$49,999	7.4	8.8	1.5
\$50,000 - \$59,999	4.6	5.9	1.2
\$60,000 - \$79,999	3.8	4.6	0.8
\$80,000 - \$99,999	1.1	1.3	0.2
\$100,000 and over	1.3	1.6	0.4
Don't know/not stated*		26.3	
Dissimilarity Index			7.8
Household income:			
No income or loss	0.3	0.5	0.2
Under \$10,000	5.7	3.8	1.9
\$ 10,000 - \$ 19,999	11.1	13.1	2
\$ 20,000 - \$ 29,999	11.8	14.7	2.9
\$ 30,000 - \$ 39,999	11.8	14.2	2.4
\$ 40,000 - \$ 49,999	11.5	14	2.4
\$ 50,000 - \$ 59,999	10.7	11.7	1
\$ 60,000 - \$ 79,999	16.5	13	3.4
\$ 80,000 - \$ 99,999	9.6	7.6	2
\$100,000 and over	11.1	7.4	3.6
Don't know/not stated*		33.8	
Dissimilarity Index			11

Notes:

*Not stated values are shown for information purposes only. They are not included in the estimations of the distributions.

**Combines the categories French and other, and English, French and other.

(1) Census universe = employed labour force including absent; GSS universe = persons who worked in week prior to survey excluding absent.

The unusually low value of the DI for unpaid care given to seniors reflects, in part, the small number of possible response categories. Also, these data might be more useful as indicators of the census-GSS divergence, if they are limited to people in the prime ages where care for seniors is likely to be an issue in their lives. However, as Table 2.2 shows, the picture changes little when the data are restricted to respondents aged 45 to 64.

As regards the apparent direction of bias in the census distributions, there is a consistent tendency among all three unpaid work variables for the census distributions to be more heavily weighted at the lower tails of the distribution (the zero-hours category). At the upper

tails, an apparent "overweighting" of the census distributions is seen for housework, and among men only for child-care work.

Table 2.2 Population Aged 15 and Over, Hours of Unpaid Care to Seniors, by Age and Sex

	Total			Age 25-44		
	Census	GSS	Difference	Census	GSS	Difference
Male						
None	86.4	84.0	2.4	87.0	84.0	3.0
Less than 5 hours	9.6	10.1	0.5	9.4	10.6	1.2
5 to 9 hours	2.4	3.4	1.0	2.2	3.2	1.0
10 or more hours	1.6	2.5	0.9	1.4	2.1	0.7
Not stated		3.0			3.0	
Dissimilarity Index			2.4			3.0
Age 45-64			Age 65 and over			
Male						
None	82.5	82.0	0.6	86.4	82.1	4.4
Less than 5 hours	12.1	10.2	1.9	7.8	10.7	2.9
5 to 9 hours	3.3	4.6	1.4	2.9	3.4	0.5
10 or more hours	2.1	3.2	1.1	2.9	3.9	1.0
Not stated		3.1			4.4	
Dissimilarity Index			2.4			4.4

	Total			Age 25-44		
	Census	GSS	Difference	Census	GSS	Difference
Female						
None	80.8	78.4	2.4	80.6	79.6	1.0
Less than 5 hours	12.0	12.9	0.9	12.9	12.2	0.7
5 to 9 hours	4.1	4.0	0.1	3.9	3.8	0.1
10 or more hours	3.1	4.7	1.6	2.6	4.4	1.8
Not stated		2.7			2.1	
Dissimilarity Index			2.5			1.8
Age 45-64			Age 65 and over			
Female						
None	75.0	72.1	2.9	83.4	79.1	4.2
Less than 5 hours	14.4	16.3	1.9	8.8	11.1	2.4
5 to 9 hours	5.9	5.0	0.9	4.0	4.5	0.5
10 or more hours	4.7	6.6	1.9	3.9	5.2	1.3
Not stated		3.0			4.7	
Dissimilarity Index			3.8			4.2

Note:

Due to small cell sizes in the GSS, data for respondents aged 15 to 24 are not shown.

Population Subgroups with Unusually High Levels of Census-GSS Divergence

Housework

Much above average levels are seen for the youngest and the oldest age groups (15 to 24 and 65+, respectively). Here, we must note the high levels of non-response in the GSS data for the 65+ age group: 7% for men and 10% for women. As well, those who lived alone, or did not work in the week before the survey, or had a non-English home language also had above average DI values.

Table 2.3 shows that, among men, the DI rises above the average only for men living alone. The DI is below average for men and women who have a spouse and a child in their homes. Among women, the DI is above average for those living alone and for lone-parent mothers with a child under 15 in the home. These observations seem inconsistent with the hypothesis that a higher level of proxy reporting in the census was damaging to its quality, compared to that of the GSS.

The DI values for both men and women who had a paid job in the week before the survey are substantially below the average and far below those for persons who did not have such a job. The latter difference is so great as to suggest that those analyses in which one can omit persons without a paid job in the week before the survey could have census data with quality virtually on par with that of the GSS.

Table 2.3 Population in Private Households, Number of Hours of Housework, by Sex and Living Arrangements

Hours of unpaid housework	Census	GSS	Difference	Census	GSS	Difference	Census	GSS	Difference
	Total			Living alone			Spouse only		
				Male					
None	15.4	10.1	5.3	15.0	7.4	7.6	13.0	11.3	1.7
Less than 5 hours	30.1	29.9	0.2	29.6	36.3	6.7	24.0	25.8	1.8
5 to 14 hours	32.8	40.5	7.7	36.5	43.6	7.1	34.5	40.8	6.3
15 to 29 hours	14.3	14.9	0.6	13.4	10.4	3.0	17.6	16.6	1.0
30 or more hours	7.4	4.6	2.8	5.5	2.3	3.2	10.9	5.5	5.4
Not stated		3.9			5.0			3.8	
Dissimilarity Index			8.3			13.8			8.1
	Spouse with at least 1 child under the age of 15			Spouse with youngest child 15 and over					
				Male					
None	7.9	6.4	1.5	13.4	16.4	3.0			
Less than 5 hours	24.6	21.2	3.4	26.3	21.4	4.9			
5 to 14 hours	39.4	45.6	6.2	35.9	38.2	2.3			
15 to 29 hours	19.1	20.0	0.9	16.2	17.9	1.7			
30 or more hours	9.1	6.8	2.3	8.2	6.2	2.0			
Not stated		4.3			3.1				
Dissimilarity Index			7.1			6.9			

Table 2.3 (Continued)

Hours of unpaid housework	Census	GSS	Difference	Census	GSS	Difference	Census	GSS	Difference
	Total			Spouse only			Spouse with at least 1 child under the age of 15		
Female									
None	7.7	4.3	3.4	5.4	4.7	0.7	1.8	2.4	0.6
Less than 5 hours	15.7	16.1	0.4	9.9	12.0	2.1	4.7	5.7	1.0
5 to 14 hours	28.1	35.3	7.2	31.1	39.5	8.4	22.5	27.3	4.8
15 to 29 hours	24.0	27.0	3.0	27.6	29.8	2.2	30.4	34.2	3.8
30 to 59 hours	17.0	14.4	2.6	20.3	11.4	8.9	24.9	24.8	0.1
60 or more hours	7.6	2.8	4.8	5.7	2.7	3.0	15.6	5.7	9.9
Not stated		4.2			5.3			3.1	
Dissimilarity Index			10.7			12.6			10.1
	Living alone			Spouse with youngest child 15 and over			Lone parent with youngest child 15 and over		
Female									
None	12.9	6.6	6.3	3.6	3.7	0.1	7.0	3.4	3.6
Less than 5 hours	19.8	21.0	1.2	6.3	10.0	3.7	10.5	14.1	3.6
5 to 14 hours	35.5	45.4	9.9	25.5	21.8	3.7	30.5	36.1	5.6
15 to 29 hours	19.8	20.1	0.3	30.9	37.5	6.6	27.3	28.3	1.0
30 or more hours	12.0	6.9	5.1	33.7	27.0	6.7	24.7	18.1	6.6
Not stated		7.4			3.6			4.4	
Dissimilarity Index			11.4			10.4			10.2
	Lone parent with at least 1 child under the age of 15								
Female									
None	2.6	1.1	1.5						
Less than 5 hours	6.3	8.9	2.6						
5 to 14 hours	26.2	34.2	8.0						
15 to 29 hours	29.2	32.2	3.0						
30 to 59 hours	21.4	18.6	2.8						
60 or more hours	14.3	5.0	9.3						
Not stated		2.5							
Dissimilarity Index			13.6						

Note:

Due to small cell sizes in the GSS, data for male lone parents could not be compared. The categories 30-59 hours and 60 or more hours are collapsed for women living alone, living with a spouse and one child aged 15 and over, and lone parents living with a child aged 15 and over.

This is particularly notable regarding men, since their propensity to have had such a job would be generally very high across the country. This proposition, if correct, would be yet another reason for suggesting that it is quite wrong to allege that the census data for men are especially weak compared to those of the GSS. Also notable are the low DI values for both men and women who have a university education. This suggests that where an analysis can be restricted to university-educated persons with a job in the week before the survey, the quality of the census data is likely to be on a par with that of the GSS.

Child Care

Patterns, broadly similar to those just cited for housework, are seen in the DI data for child care. One notable difference applies to lone-parent mothers, where the census-GSS discrepancy is below average in the case of child care.

The child-care data are limited to persons who had a child aged less than 15 in their homes, and are especially worthy of study for the 25 to 44 age group (Table 2.4). In this age group, the DI for men is 15%, two points above the average. The DI for women is at the average of 13%. For both sexes, the DI level is being generated largely by apparent "overweighting" of the census distributions in the lower section of the scale of hours spent doing child care.

This suggests that the census data are not overstating men's or women's contribution to child-care work. The greater discrepancy for men may have little to do with proxy reporting (if we assume the wife is filling out the questionnaire) since both sexes show marked "overweighting" of the census distributions in the lower section of the scale.

**Table 2.4 Population Aged 15 and Over, Hours of Unpaid Child Care,
by Sex and Age Groups**
(Households with at least one child less than 15)

	Total			Age 25-44			Age 45 and over		
	Census	GSS	Difference	Census	GSS	Difference	Census	GSS	Difference
							Male		
None	23.1	16.8	6.3	12.0	6.0	6.1	25.5	22.3	3.1
Less than 5 hrs	17.8	14.0	3.8	15.9	10.8	5.0	23.5	17.5	6.0
5 to 14 hours	25.6	24.7	0.9	29.5	27.8	1.7	27.0	24.8	2.2
15 to 29 hours	16.7	25.1	8.4	21.1	30.5	9.4	12.5	22.0	9.5
30 to 59 hours	9.3	13.5	4.2	12.1	17.7	5.6	11.5	13.4	1.9
60 or more hours	7.5	6.0	1.5	9.4	7.2	2.2			
Not stated		3.8			4.3			5.1	
Dissimilarity Index			12.6			15.0			11.4
Female									
None	14.8	9.9	4.8	6.6	2.6	4.0	22.0	17.2	4.9
Less than 5 hours	8.2	5.3	2.8	5.1	2.2	3.0	13.6	11.6	2.1
5 to 14 hours	16.1	12.2	4.0	16.2	11.2	5.1	24.1	24.6	0.6
15 to 29 hours	16.4	15.4	1.0	18.9	18.7	0.2	16.6	9.1	7.5
30 to 59 hours	17.3	24.5	7.2	20.9	26.6	5.7	23.7	37.6	13.9
60 or more hours	27.2	32.6	5.4	32.2	38.8	6.6			
Not stated		2.8			2.9			4.8	
Dissimilarity Index			12.6			12.3			14.5

Note:

Due to small cell sizes in the GSS, data for respondents aged 15 to 24 are not shown. Data for those aged 45 to 64 and those 65 and over have been collapsed. The categories 30 to 59 hours and 60 hours or more have been collapsed for men and women aged 45 and over.

For both men and women, the subset living with a spouse, has DI values right at the average of 13%. This implies that the 15% value cited above for men is above the average because of men in other living arrangements.

In sum, the DI levels for child care are generally close to the danger zone of 15% or more. However, they are not markedly worse for men than for women where we restrict observations for persons living with a spouse/partner. They are better than the overall average (for child care) in the case of lone-parent mothers. The latter population's 10% DI is well within the range established for census-GSS divergence across a representative selection of census variables.

Senior Care

Patterns, broadly similar to those cited above for housework, are also seen in the DI data for senior care. In particular, higher than average DI values are observed for men living with a spouse only, lone-parent mothers, persons who did not work in the week before the survey or those with a non-English home language. However, the levels of the DI regarding senior care are so low (almost always below 5%, see Table 2.5) that further discussion for this variable seems unwarranted. The data suggest that, with regards to senior care, the census data are on a par with the GSS as far as quality is concerned.

Table 2.5 Population in Private Households, Hours of Unpaid Care to Seniors, by Living Arrangement

	Total			Living alone			Spouse only		
	Census	GSS	Difference	Census	GSS	Difference	Census	GSS	Difference
	Male								
None	86.4	84.0	2.4	89.3	85.9	3.4	86.6	81.1	5.5
Less than 5 hours	9.6	10.1	0.5	7.1	8.6	1.5	9.0	11.9	2.9
5 to 9 hours	2.4	3.4	1.0	2.2	2.4	0.3	2.6	3.7	1.1
10 or more hours	1.7	2.5	0.9	1.5	3.1	1.7	1.8	3.3	1.5
Not stated		3.0			3.3			3.0	
Dissimilarity Index			2.4			3.4			5.5
	Spouse with at least 1 child under the age of 15			Spouse with youngest child 15 and over					
	Census	GSS	Difference	Census	GSS	Difference			
	Male								
None	83.4	83.4	0.0	81.9	81.6	0.2			
Less than 5 hours	12.3	10.2	2.2	13.1	13.2	0.1			
5 to 9 hours	2.8	4.2	1.4	5.0	5.1	0.1			
10 or more hours	1.5	2.3	0.8						
Not stated		3.4			2.6				
Dissimilarity Index			2.2			0.2			

Table 2.5 (Continued)

	Total			Living alone			Spouse only		
	Census	GSS	Difference	Census	GSS	Difference	Census	GSS	Difference
	Female								
None	80.9	78.4	2.4	83.8	79.2	4.6	80.9	78.3	2.6
Less than 5 hours	12.0	12.9	0.9	10.0	11.4	1.3	11.1	13.0	1.8
5 to 9 hours	4.1	4.0	0.1	3.9	4.8	0.9	4.4	3.9	0.5
10 or more hours	3.1	4.7	1.6	2.3	4.7	2.4	3.5	4.8	1.3
Not stated		2.7			3.9			3.5	
Dissimilarity Index						4.6			3.1
	Spouse with at least 1 child under the age of 15			Spouse with youngest child 15 and over			Lone parent with at least 1 child under the age of 15		
	Census	GSS	Difference	Census	GSS	Difference	Census	GSS	Difference
	Female								
None	77.8	80.7	2.9	73.6	71.1	2.6	81.0	77.7	3.4
Less than 5 hours	14.9	12.6	2.4	16.1	16.7	0.6	11.7	9.9	1.8
5 to 9 hours	4.5	2.7	1.8	5.9	5.1	0.8	4.3	6.8	2.5
10 or more hours	2.8	4.0	1.2	4.4	7.2	2.8	3.0	5.6	2.6
Not stated		2.6			1.6			1.9	
Dissimilarity Index			4.1			3.4			5.1

Note:

Due to small cell sizes in the GSS, data for male lone parents and female lone parents with a child 15 and over could not be compared.

As we close this discussion and get ready to develop an evaluation using the GSS time-use data (where the strength of the GSS is considered to be transparently pre-eminent), it is important to observe that nothing similar to the census data on senior care is available in the GSS time-use data. There is no way to extract useful estimates of time spent doing senior care from the 1986 or 1992 time-use surveys (GSS). Hence, for this kind of care, there is no contest—the census data are the winners.

To be sure, there are serious limitations in the census senior care data, arising from the fact that the census does not allow one to determine which respondents had senior relatives who might need care. However, by limiting analysis to situations where the respondent has a senior in the home, especially where the senior is reportedly disabled, it becomes possible to make useful applications of the census senior care data that are utterly impossible with the GSS. (For a related discussion see Stone and Silver, 1998a).

Conclusion

In conclusion, only the data for child care show levels of the DI that are close to a danger zone where there are signals that unacceptably weak data may exist. Even the child-care data have DI values within the range of values observable across the spectrum of those census variables whose legitimacy for inclusion in the census is not questioned.

Moreover, if the analysis can be restricted to those who had a job or a university-level education, one has enhanced the quality of the census data subset being used, at least in comparison with using the entire census data set.

Indeed, with the data on housework, that restriction places us in a situation where it is impossible to say whether the census or the GSS has better data quality, as far as the stylized time-expenditure questions are concerned. This perhaps controversial remark applies to all the data on senior care. The situation concerning senior care is especially important because the GSS time-diary data (where the case favouring the GSS is most strongly made) offer no opportunity to provide estimates of time spent caring for seniors.

Finally, contrary to our expectations, there are no indications in this study that the census unpaid work data for men are of weaker quality than those for women.

In short, the quality-based case against the census unpaid work data cannot be made on the basis of what we know about non-response patterns or the prevalence of inappropriate multiple responses (the subject of our first paper). This (second) paper demonstrates that a case can also not be made on the basis of the stylized time-use question, despite the fact that both surveys used identical questions. Thus, the quality-based case against the census unpaid work data will now have to rest 100% on what turns up in our analysis of the 1992 GSS time-diary data. (Though the 1998 GSS time-diary data are a better choice, they are not yet available.) This will be the subject of our next paper.

PAPER 3: QUALITY OF THE 1996 CENSUS UNPAID WORK DATA IN LIGHT OF COMPARISONS WITH 1992 GSS BASED ON DIARY DATA

Introduction

This paper evaluates the quality of the results of the 1996 Census questions on unpaid housework and unpaid child care, by means of comparisons with data yielded by the time-diary data from the 1992 General Social Survey (GSS). This study tests whether certain patterns of difference between diary and stylized data, as described by the literature, are found in reasonably comparable data for selected distributions from the 1992 GSS and the 1996 Census.

Only the census questions on housework and child-care activity are addressed in this paper, as the 1992 GSS diary data are not suitable for an analysis of time spent on care to seniors. The 1998 GSS has attempted to address this defect with special probing concerning care given to seniors.

This is the third and final paper in a series of reports on the evaluation of the quality of the census questions on unpaid work. The first paper looked at the quality of the census data in terms of non-response rates and levels of imputation caused by manifestly inappropriate responses (Swain and Stone, 1998). The second paper examined the quality implications of differences between distributions from the 1996 Census and corresponding distributions from the 1996 GSS, given that both surveys used identical questions (Stone and Swain, 1999). A fourth paper on potential analytical uses of the census data is also available (Stone and Silver, 1998a).

Methodology

Because of the limited comparisons that can actually be made between the diary and stylized data (see below), the hypothesis to be tested in this paper is very simple. Generally, we expect that when time use estimates obtained using the diary method are compared with estimates obtained using stylized questions, the patterns of the distribution for housework should be reasonably similar. In the case of child care, the stylized method should produce much higher estimates than the diary method where only primary child-care time is counted.

Again, because of the basic non-comparability of the GSS diary data and census data, no actual numbers are compared. Rather, patterns in the data are conveyed through a series of charts.

Comparability of the Census and the GSS

Many of the sources of non-comparability between the 1996 GSS and the census outlined in the second paper apply also to the 1992 GSS, namely, the telephone/interviewer method of data collection, differences in reference period, proxy reporting and handling of non-response. (For a full discussion of these differences see Stone and Swain, 1999.) With the introduction of

the time-diary data, however, new sources of non-comparability between the census and the GSS are added to those already discussed.

As mentioned in the second paper, one source of non-comparability between the census and the GSS was the reference period. As was the case with the 1996 GSS, 1992 GSS data were collected over a year, whereas the census data were collected at only one point in time. In addition to this seasonal variability in the data, there is now also the basic difference of the year in which the surveys took place: 1992 versus 1996.

A most important and fundamental difference between the census and 1992 GSS time-diary data is how the information on time spent on unpaid work activities is collected and estimated. As the time diary suggests, respondents are asked to report all activities carried out over a 24-hour period—usually the day before the interview—in a diary fashion. The activities are then coded according to a pre-established set of codes. To arrive at an estimate of the amount of time spent on an activity such as housework, the times reported for individual activities, such as meal preparation, laundry and cleaning, are added together. Thus, it is the researcher who defines which activities constitute unpaid household or caring work.

In contrast, the census used three stylized questions to collect information on the three separate unpaid work activities. The respondent provided an estimate of time spent doing each of the activities for the seven-day period preceding the census. To reduce respondent burden, the census questions used answer categories to record the time spent rather than exact estimates. Thus, it was the respondent who ultimately determined which activities were included under each of the unpaid activity questions. This can result in a wide range of interpretation by respondents, particularly the housework question.

In short, so great is the collection of non-comparabilities between the census and the GSS time-diary data that it is best to adopt, as a working hypothesis, the notion that the two sources are measuring different variables. Each measured variable is useful for a specific class of analytical questions. The different classes or analytical questions may overlap, but they are not equivalent.

Stylized versus Time-Diary Methods of Data Collection

In his review of the literature on diary and stylized methods of data collection, Paillé (1994) found that the accuracy of estimates from the diary method is generally believed to be superior to that from the stylized approach. The diary method uses a shorter recall time (usually the previous 24 hours). Therefore, activities are more easily remembered. Respondents merely report their activities in chronological order to be coded later by the researcher.

On the other hand, because diaries collect information for only one day in each respondent's life, infrequent activities can easily be underestimated. Volunteer work is an activity that often takes place in a few days of each week or each month for a given respondent and, thus,

its prevalence in the population is probably underestimated in diary data (for a related discussion see Paillé, 1994).

The GSS diaries usually collect information on so-called primary activities only. They do not collect information on other activities that occur simultaneously, a most important example of which is child care.

Stylized questions generally use a longer reference period, such as a week or a month, which can lead to recall difficulties for the respondent. The respondent must also decide which activities to include when deciding how much time has been spent on an activity such as housework.⁶

Finally, in comparison with diary data, data from stylized questions may overestimate time spent on activities that often occur simultaneously with other activities. This is because respondents are asked to estimate time spent on each activity separately, so when all time reported in each question is added together, the total time estimated is often larger than the reference period.

However, that property of data based on stylized questions is not a weakness where it is important that multitasking within a specific time interval be considered. In this situation, activities frequently done in multitasking mode may be better measured with stylized questions than with a time-use diary that restricts every moment of time to one and only one primary activity. (This restriction is not intrinsic to time-use diary data, but it is a property of the GSS diaries.)

In this context, there is no overestimation arising from the stylized questions. Analysts simply need to know how to exercise due caution in adding numbers across activities. This is a minor problem because data based on stylized questions are not often used to arrive at estimates of total time spent doing unpaid work.

This issue of measurement of multiple activities within a given time period is of fundamental concern in the estimation of the time spent on child care. Most experts on child-care work believe that total time spent in the care of children cannot be adequately appreciated unless multitasking is taken into account.

Returning to the hypotheses to be tested in this paper then, the literature suggests that estimates of time spent on housework should be roughly similar between the two methods. In the case of time spent on child care, however, the stylized questions should produce larger estimates (higher proportions in the upper tails of the distribution) than the time-diary data.

Housework

The list of activities from the GSS diary data used to estimate time spent on unpaid housework is taken from the Total Work Accounts System (TWAS) data base (Stone and Chicha, 1997). It must be noted that the TWA definition of "unpaid work for self and others" is fairly broad, and it is very unlikely that census respondents would themselves have included

all the activities involved with the "unpaid work for self and others" variable of the TWAS. (For the full definition of "unpaid work for self and others" used in the TWA, see Appendix 3.A.) The figures presented below, therefore, provide only a rough indication of the comparability of the diary and stylized data on unpaid housework.

Looking at the data for women first of all (Figure 3.1), the distributions appear fairly similar except for the two-to-four hour category. However, this pattern could be the result of the differences in the list of activities that could be included as housework as mentioned above.

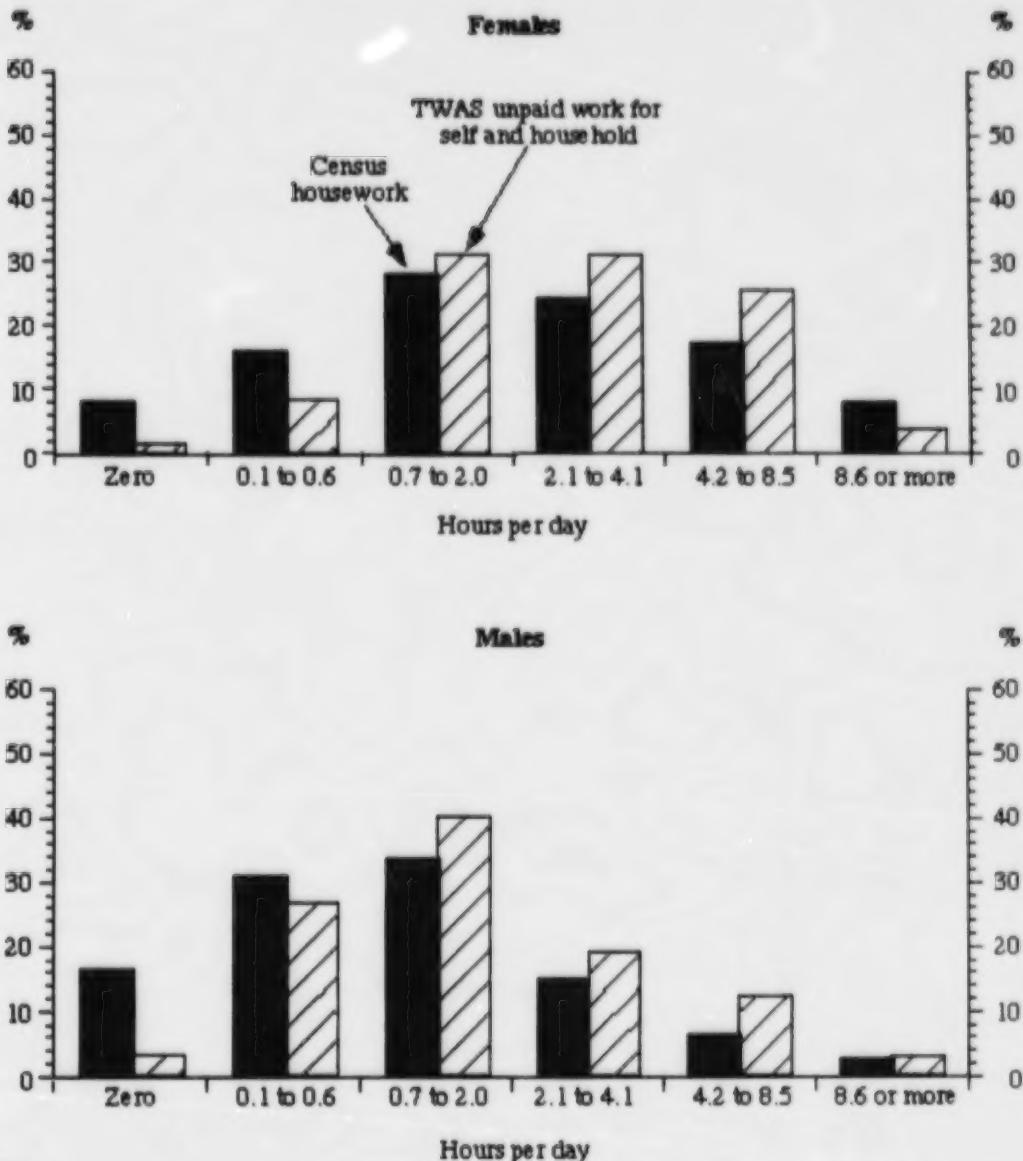
Turning to the data for men, the distributions from the two sources appear quite similar. The diary data show a higher proportion of men spending 4.3 to 8.5 hours on unpaid housework than the census. Again, this may be the result of differences in the housework definition.

On the other hand, the much higher proportion of men in the census with zero hours of housework than in the GSS may lead us to suspect that the census data for men are underreported. While there may be several reasons for this, the most often cited is proxy reporting. This would assume that most proxy reporting is done by women and that women consistently underestimate the amount of housework performed by their spouses. But men, themselves, may underestimate time spent on household work, either because they think the census question does not apply to them, or because they think the activity or time spent on it is too insignificant to count. For example, taking out the garbage or changing a light bulb takes little time and may be ignored or even forgotten over the course of a week.

Child Care

Figure 3.2 compares time spent caring for children among respondents with a child under the age of 15 in the home. Here we find our initial hypothesis clearly supported. For both men and women, the stylized question in the census has resulted in higher estimates of time spent on child care compared to the GSS diary data.

Figure 3.1. Distributions of Population by Categories of Hours Spent Doing Housework (1996 Census) and Hours Spent Doing Unpaid Work for Self and Household (1992 TWAS)¹ (Population aged 15 or more)



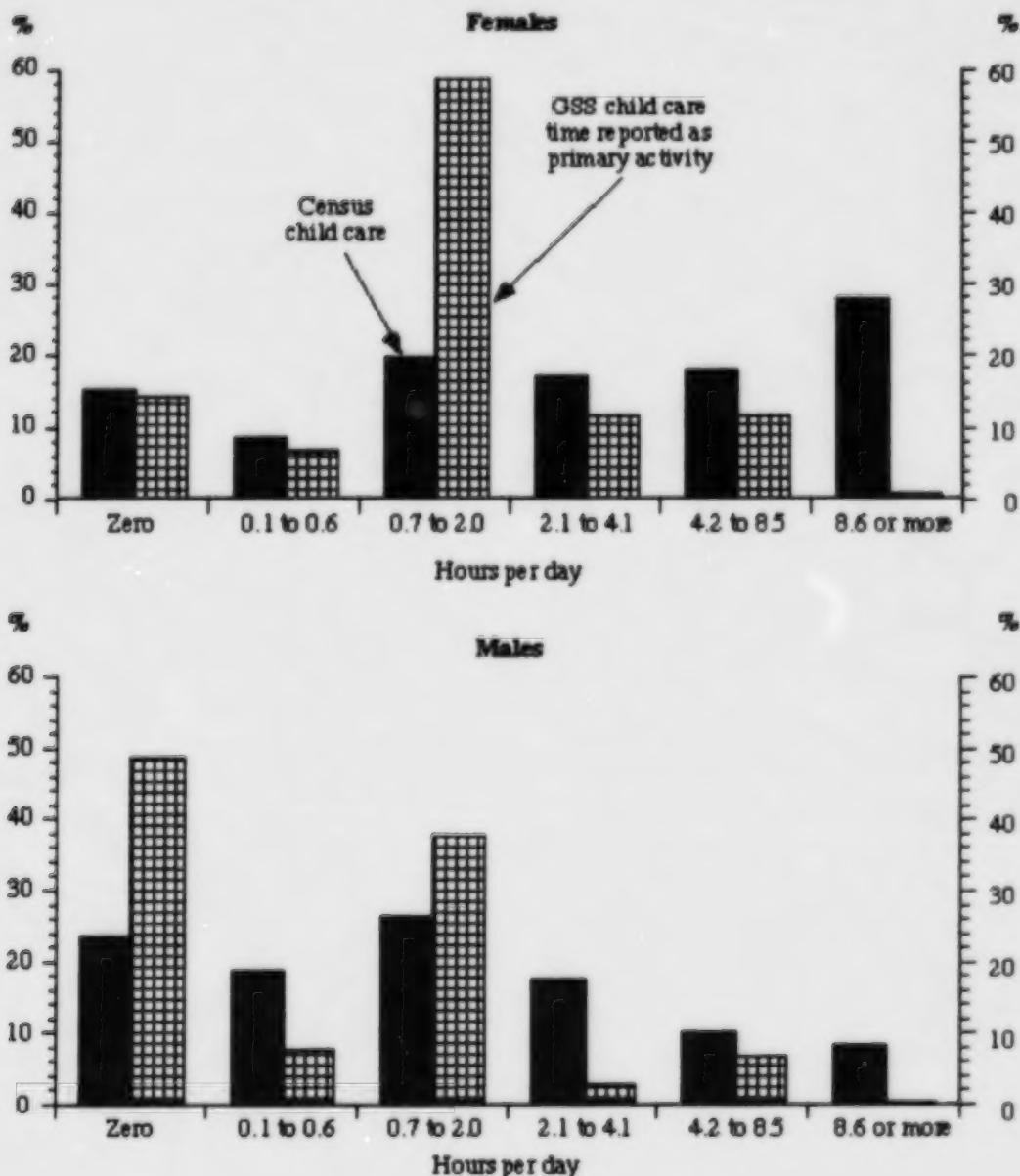
Note:

¹ Total Work Accounts System, based on the 1992 General Social Survey.

Source:

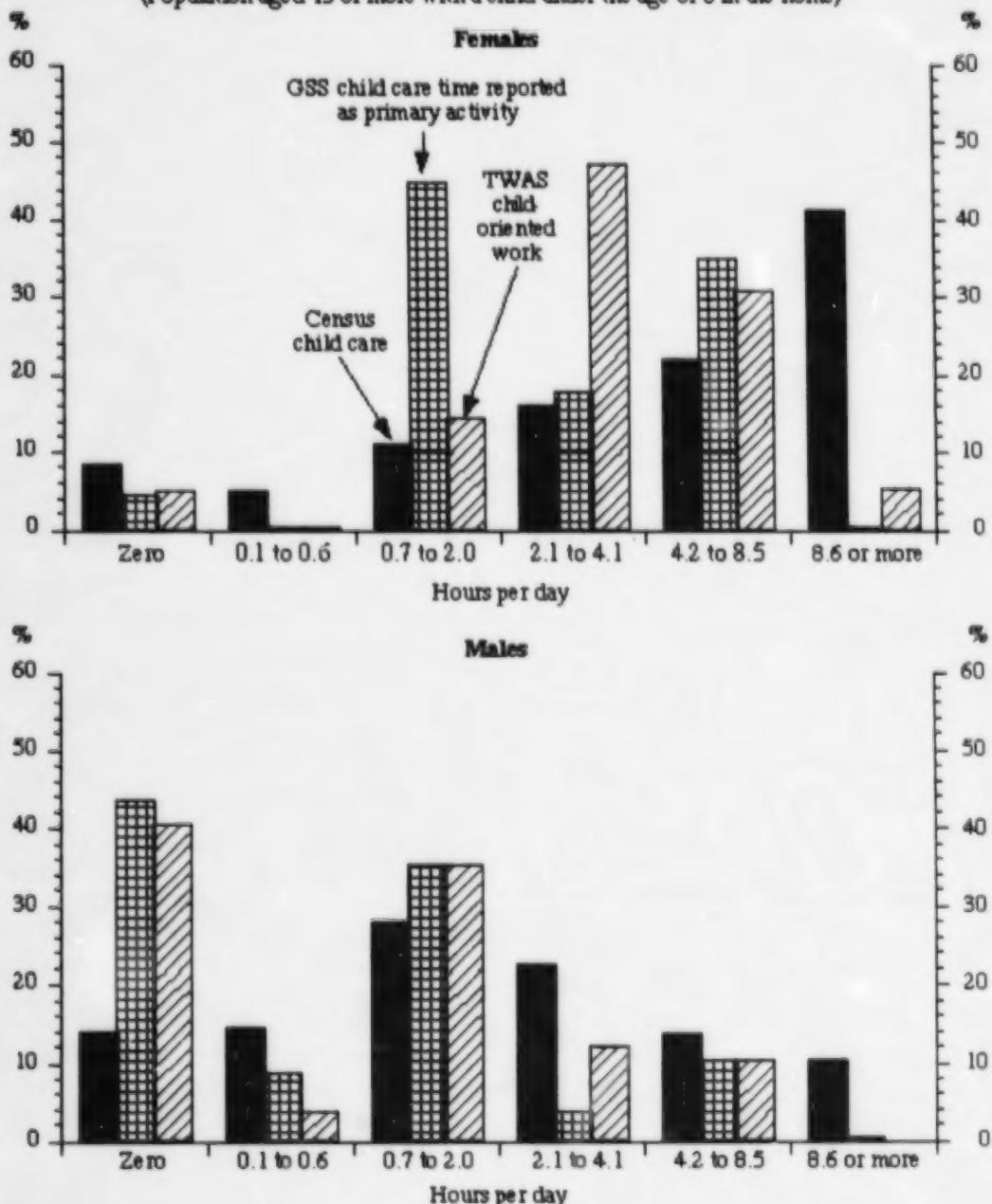
Statistics Canada, 1996 Census and the Total Work Accounts System (based on the General Social Survey).

Figure 3.2. Distributions of Population According to Categories of Hours Spent Doing Unpaid Child Care Work, by Sex, 1996 Census and 1992 General Social Survey
 (Population aged 15 or more with a child under the age of 15 in the home)



Sources:
 Statistics Canada, 1996 Census and the 1992 General Social Survey.

Figure 3.3. Distributions of Population According to Categories of Hours Spent Doing Unpaid Child Care Work, by Sex, 1996 Census and 1992 General Social Survey
 (Population aged 15 or more with a child under the age of 6 in the home)



Source:

Statistics Canada, 1996 Census, 1992 General Social Survey and the Total Work Accounts System (based on the General Social Survey).

Looking at the distributions for women, Figure 3.2 shows the GSS respondents to be heavily concentrated in the 0.7 to 2.0 hours category. Basically, the GSS data are saying that only about 20% of these women spend more than two hours per day in child-care work. While it is possible that children over five do not require more than two hours of care, where the mother's attention is 100% focussed on the child, this pattern only makes sense if we are trying to measure child care as the primary activity. But as we have argued, to have a true measure of time spent on child care, secondary activity or multitasking must also be considered. Hence, despite the weakness in the census child-care data observed in our second paper, the census distribution is probably more useful to the analysis of child-care behaviour than that obtained from the GSS diary data.

Figure 3.3 lends some support to this criticism of the GSS diary data. Here we focus on households with at least one child less than six years old. Hence, we would expect the time invested in child care would be heaviest. In addition to the census and GSS child-care variables, the TWA "child-oriented work" variable is also included. The TWA variable expands the GSS definition of child care to form a new variable called "child-oriented work," by including several additional activities such as housework or shopping if a child was also present during the activity (see definition in Appendix 3.A).

By expanding the definition in this manner, the GSS distribution shifts to the right, suggesting more caring work taking place than indicated by the data based on primary activity only. Still, even the distribution using the TWA definition falls well below that obtained from the census, particularly in the upper category, 8.6 hours or more. While it may seem that the census data are overestimates, it should be noted that in our second study it was found that the proportion of women in the census reporting 60 or more hours of child care was consistently below that observed using the 1996 GSS stylized question.

As mentioned above, the data for men also found that the census stylized question resulted in a higher estimate of time spent on child care by men than shown by the diary data.

Of particular note is the large difference between the census and the GSS in the proportion of men reporting zero hours of child care. The proportion of men in the GSS is almost double that of men in the census for this category. This could be a result of the focus on child care as the primary activity in the diary data.

Conclusion

To conclude, we find that general statements about whether the time-diary data are overall more accurate than the stylized-question data cannot be applied in this census data evaluation.

Because the census measures only three specific aspects of unpaid work, total or overall unpaid work numbers cannot be computed using census data. Rather, time spent on unpaid housework, unpaid child care and unpaid care to seniors must be studied as individual non-additive sets of activities.

Our hypothesis in this paper was that a comparison of time-diary and stylized data would show similar patterns of distribution for housework, but that the stylized questions would produce much higher estimates of child care than the diary data. The data presented in this paper have supported this hypothesis. Even the patterns for housework, although based on very divergent definitions according to the TWA and what we might expect in the census, did not show the census data to be off base when compared with the diary data.

In our analysis of the child-care data, we found that the diary data do not capture the whole story as we see it. The concept of multitasking is fundamental to the understanding of time spent caring for children. Thus, we find the census data are better than the time diary data for measuring this activity.

Finally, with respect to unpaid care for seniors, there is no possibility of reasonable analysis of this activity using time-diary data.

Appendix 3.A: Operational Definitions of Destinations of Work of Economic Value

"The following is a set of operational definitions that are strictly linked to the properties and limitations of the 1992 GSS database. The definitions often represent an effort to impute specific beneficiaries to work activities on the basis of the nature of the activity and the presence of particular persons during the conduct of the activity.

"This imputation was needed because the GSS questionnaire did not ask respondents to identify the perceived beneficiaries of their work activities. This gap is a major limitation of the GSS database for the purposes of the TWAS.

..."

"Plain language cannot render some of the following definitions clearly. This is because of complex Boolean logic structures built into some of the definitions. Hence the following set of definitions in plain language are less precise than the corresponding computer program code.

..."

"Destination 6: Child – . . . doing domestic work, or routine shopping, or Care of Children, and in contact with the child during that particular work.

..."

"Destination 9: Self and other members of the household (if any) – . . . doing either of Meal Preparation, Meal Clean-Up, Routine Shopping, Washing, Dressing, Packing, Adult Medical Care (at home), and reportedly alone during that particular work (imputation of work output of economic value – service you would have to buy if you could not do it for yourself – to self)."

Source:

Stone and Chicha (1996).

BIBLIOGRAPHY

Paillé, B. (1994). *Estimating the Volume of Unpaid Activities in Canada, 1992: An Evaluation of Data from the General Social Survey*. General Social Survey Working Paper #10, Statistics Canada.

Stone, L.O. (1999). "Computing a Household-Level Index for Time Spent Doing Child Care Work." Unpublished manuscript, 2 pages.

Stone, L.O. and M.-T. Chicha. (October 1997). *Constructing the Statistics Canada Total Work Accounts System – Principles and Problems in Allocating Unpaid Work Outputs Among Alternative Destinations Using Conventional Time Use Data*. Paper presented at the 19th Reunion of the International Association for Time Use Research, Stockholm, Sweden.

———. (1996). *The Statistics Canada Total Work Accounts System*. Statistics Canada catalogue no. 89-549-XPE. Ottawa: Minister of Industry.

Stone, L.O. and C. Silver. (1998a). "Potential Contributions of the 1996 Census to Policy Applications of Data on Unpaid Work." Unpublished manuscript.

———. (1998b). *The Development of a Measure of a Household's Total Time Spent Caring for Children Using the 1996 Census Data*. Working paper.

Stone, L.O. and S. Swain. (1999). *Quality of the 1996 Census Unpaid Work Data in the Light of Comparisons with 1996 GSS Data Based on Identical Questions*. Working paper.

Swain, S. and L. Stone. (1998). *Evaluation of the 1996 Census Questions on Unpaid Work – Report No. 1 Non-Response and Multiple Response Rate*. Working paper.

GLOSSARY

Census family

Refers to a now-married or common-law couple (with or without never-married sons and/or daughters of either or both spouses) or a lone parent of any marital status, with at least one never-married son or daughter living in the same dwelling.

Children

Refers to the never-married sons and/or daughters of now-married or common-law couples and lone parents.

Collective dwelling

Refers to a dwelling of a commercial, institutional or communal nature. Included are lodging or rooming houses, hotels, motels, tourist homes, nursing homes, hospitals, staff residences, communal quarters (military camps), work camps, jails, missions, group homes and so on.

Dwelling

Refers to a set of living quarters in which a person or a group of persons reside or could reside.

Economic family

Refers to a group of two or more persons who live in the same dwelling and are related to each other by blood, marriage, common law or adoption.

Presence of children

Classifies persons 15 years of age and over living in census families into those with no children at home and those with children at home. Those persons with children at home are further classified on the basis of the age groups of all their children.

Private dwelling

Refers to a separate set of living quarters with a private entrance either from outside or from a common hall, lobby, vestibule or stairway inside the building.

ENDNOTES

¹ It is now generally recognized among analysts of the data that, with regard to child caring work at least, key 1996 Census answers categories (see Appendix A) are much too broad, and will need to be refined if such data are collected in future censuses.

² The phrase "2B questionnaire" refers to the questionnaire that had the largest number of questions and which was designed to be completed by only 20% of all census respondents. While the 2B questionnaire was the main instrument used to collect census information, the "2D canvasser questionnaire" was used in remote and northern areas. This questionnaire differed from the 2B in that some of the examples for the unpaid work questions were changed to make them more relevant to living conditions in the north. As well, this questionnaire was completed by an interviewer, whereas the 2B questionnaire is completed by the respondent. Less than 2% of households in Canada received a 2D questionnaire.

³ As noted earlier, the 2B questionnaire is distributed to a sample of one in five households in Canada. The respondents in this sample are weighted to provide estimates for the entire Canadian population. Unweighted rates provide estimates only for those respondents who actually received the 2B questionnaire.

⁴ In order to contain the time and cost of searching for a donor, the respondent records are normally divided into strata which are themselves broadly defined matching criteria, for example, females aged 25 to 44. Both the error records and their potential donors must come from the same strata; a donor cannot come from a different stratum than the record in error. In addition to the search limitations imposed by the strata, a physical limitation exists as the census data base is divided into five smaller bases for processing purposes. The five bases are East, Quebec, Ontario, West and Overseas Households. Therefore, as a minimum, the error record and donor record will match on the broad criteria defined for the strata and the geographic location.

⁵ These figures exclude the results for the overseas households. This is a small data base made up of a diverse range of respondents in diplomatic and military situations and includes persons working on vessels. Consequently, the match rate for perfect donors for these persons is lower at 73% for the housework and child care variables and 90% for the care to seniors variable. There were, however, no records for which a donor could not be found. The potential for the overseas records to affect the results of almost all analytical uses of the data is minimal.

⁶ The fact the one source uses a fixed set of activities to mean "housework" and another uses a variable set probably means the estimates from these sources are intrinsically non-comparable.

Étude d'évaluation des données du recensement de 1996 relatives au travail non rémunéré

Leroy O. Stone

et

Sandra Swain

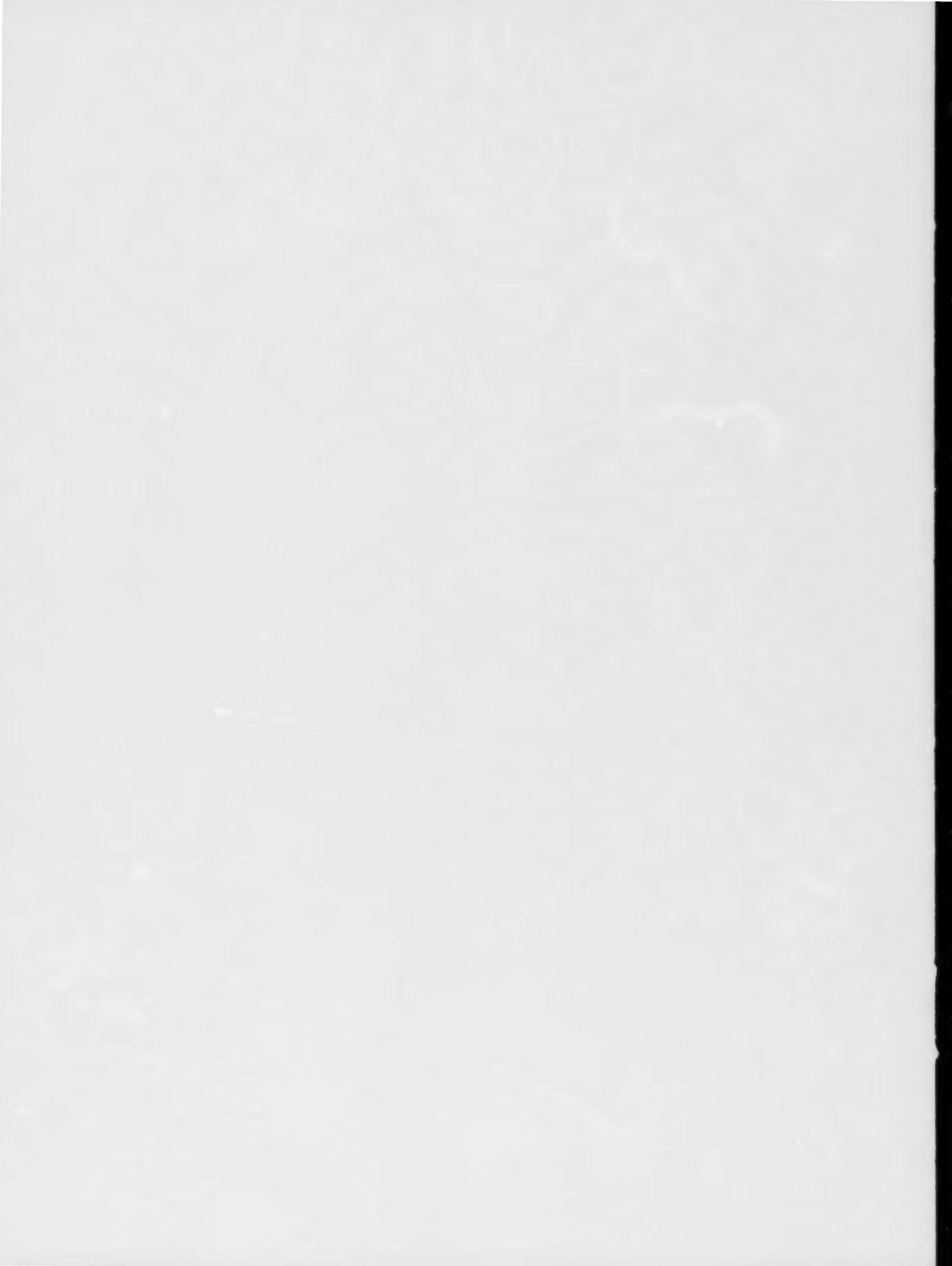
Statistique Canada



Condition féminine
Canada

Status of Women
Canada

Canada



**Étude d'évaluation des données du recensement de 1996
relatives au travail non rémunéré**

Leroy O. Stone
et
Sandra Swain
Statistique Canada

La recherche et la publication de la présente étude ont été financées par le Fonds de recherche en matière de politiques de Condition féminine Canada. Ce document exprime les opinions des auteures, qui ne correspondent pas nécessairement à la politique officielle de Condition féminine Canada ou du gouvernement du Canada.

Mars 2000

Condition féminine Canada s'est engagé à s'assurer que toute la recherche menée grâce à son Fonds de recherche en matière de politiques adhère à des principes professionnels, déontologiques et méthodologiques de haut niveau. La recherche se doit aussi d'apporter une contribution experte et unique en son genre au débat actuel sur les politiques, et d'être utile aux stratégistes de politiques, à celles et ceux qui font de la recherche, aux groupes de femmes, aux collectivités et à toute autre personne intéressée au domaine des politiques. Chaque document a été révisé anonymement par des spécialistes de la question, à qui on a demandé de faire des commentaires sur les aspects suivants :

- précision, fini et pertinence de l'information présentée;
- mesure dans laquelle l'analyse et les recommandations sont étayées par l'approche méthodologique et les données recueillies;
- contribution originale que le rapport permet d'apporter aux travaux déjà effectués dans le domaine et utilité du rapport pour les organismes oeuvrant pour l'égalité entre les sexes, les groupes de défense des droits, les stratégistes de politiques des gouvernements, les chercheuses et chercheurs et d'autre publics cibles.

Condition féminine Canada remercie toutes les personnes qui ont participé à ce processus de révision entre pairs.

Données de catalogage avant publication (Canada)

Stone, Leroy O., 1936-

Étude d'évaluation des données du recensement de 1996 relatives au travail non rémunéré

Texte en français et en anglais disposé tête-bêche.

Titre de la p. e t. addit. : The 1996 unpaid work data evaluation study

Comprend des références bibliographiques.

Également publ. sous forme électronique sur le réseau informatique. Internet.

ISBN 0-662-83874-9

N° de cat. SW21-40/2000F

1. Économie domestique — Canada — Statistiques — Évaluation.
2. Femmes — Canada — Conditions économiques — Statistiques — Évaluation.
3. Division sexuelle du travail — Canada — Statistiques — Évaluation.
4. Canada, Recensement, 1996.
5. Enquêtes sociales — Canada — Évaluation.

I. Swain, Sandra Leigh.
II. Canada. Condition féminine Canada.
III. Titre.

HQ1381.5S76 2000 339.5'088'649'0971 C99-980329-8

Gestion du projet : Julie Dompierre, Condition féminine Canada

Coordination de l'édition : Mary Trafford, Condition féminine Canada

Révision : PMF Editorial Services Inc.

Traduction : Mayn Berbery

Pour d'autres renseignements, veuillez communiquer avec la :

Direction de la recherche

Condition féminine Canada

350, rue Albert, 5^e étage

Ottawa (Ontario) K1A 1C3

Téléphone : (613) 995-7835

Télécopieur : (613) 957-3359

ATMÉ : (613) 996-1322

Courriel : research@swc-cfc.gc.ca

Ce document est aussi accessible sur le site Web de Condition féminine Canada, à l'adresse <http://www.swc-cfc.gc.ca/>.

TABLE DES MATIÈRES

PRÉFACE	iii
REMERCIEMENTS	v
IMPORTANCE DES DONNÉES DU RECENSEMENT SUR LE TRAVAIL NON RÉMUNÉRÉ ET PRINCIPALES QUESTIONS ABORDÉES DANS L'ÉTUDE D'ÉVALUATION	1
Introduction.....	1
Principales questions abordées	1
Résultats choisis	2
Principales observations relatives aux résultats.....	3
ARTICLE 1 : NON-RÉPONSE ET IMPUTATION DANS LES CAS DE RÉPONSES MULTIPLES	6
Introduction.....	6
Contexte.....	6
Qualité des données	6
Taux de non-réponse	7
Taux de réponses multiples	11
Vérification et imputation	12
Imputation dans les cas de non-réponse	14
Résolution des réponses multiples erronées	15
Conclusion	15
Annexe 1.A : Questions du recensement sur le travail non rémunéré et instructions du guide.....	17
ARTICLE 2 : QUALITÉ DES DONNÉES DU RECENSEMENT DE 1996 SUR LE TRAVAIL NON RÉMUNÉRÉ À LA LUMIÈRE DE COMPARAISONS AVEC LES DONNÉES DE L'ESG DE 1996 FONDÉES SUR DES QUESTIONS IDENTIQUES	20
Introduction.....	20
Méthodologie	20
Comparabilité des données du recensement et de l'ESG	22
Écarts globaux.....	25
Sous-groupes de la population caractérisés par des niveaux anormalement élevés de divergence entre le recensement et l'ESG	30
Conclusion	37
ARTICLE 3 : QUALITÉ DES DONNÉES DU RECENSEMENT DE 1996 SUR LE TRAVAIL NON RÉMUNÉRÉ À LA LUMIÈRE DE COMPARAISONS AVEC LES DONNÉES RELATIVES À L'EMPLOI DU TEMPS TIRÉES DE L'ESG DE 1992	39

Introduction	39
Méthodologie	39
Comparabilité des données du recensement et de l'ESG	40
Méthodes de collecte de données : questions directes et journaux	
de l'emploi du temps.....	41
Conclusion	48
Annexe 3.A : Les définitions opérationnelles du travail ayant	
une valeur économique	49
RÉFÉRENCES	50
GLOSSAIRE	51
NOTES.....	52

PRÉFACE

La décision de produire les *Études d'évaluation des données du recensement de 1996 relatives au travail non rémunéré* a été prise dans la foulée de discussions entre Condition féminine Canada et le Comité externe du Fonds de recherche en matière de politiques, au printemps de 1998. Le Comité ayant déterminé que la prestation de soins non rémunérés était un enjeu urgent qui nécessitait une attention immédiate, Statistique Canada a été invité à procéder à une analyse préliminaire des données du recensement du Canada de 1996, lequel comprenait, pour la première fois, trois questions sur le travail non rémunéré.

Les *Études d'évaluation des données du recensement de 1996 relatives au travail non rémunéré* sont une série de trois ouvrages. Le présent document, le premier des trois, a pour but d'examiner les principaux aspects de la *qualité* des données du recensement sur le travail non rémunéré. Des aspects de l'engagement paternel et des soins aux personnes âgées feront l'objet de deux autres études analytiques que Condition féminine Canada pense publier et qui reposeront aussi sur les données du recensement de 1996 concernant le travail non rémunéré. Des renseignements additionnels sur les tendances relatives au partage des soins dispensés aux enfants par les membres du ménage autres que le père et la mère accompagneront le document sur l'engagement paternel, le recensement de 1996 ayant recueilli des données pertinentes auprès de tous les membres (âgés de 15 ans et plus) des ménages de l'échantillonnage – ce qui représente une innovation importante.

Les données sur le travail non rémunéré peuvent être utilisées pour obtenir des résultats de recherche applicables à un large éventail de secteurs politiques, notamment :

- l'examen des formules de garde d'enfants dont on dispose pour répondre aux besoins actuels;
- la planification et la prestation d'une aide destinée aux personnes qui consacrent une bonne partie de leur temps à dispenser des services personnels à des personnes âgées qui en ont besoin;
- l'établissement de liens entre les services hospitaliers, les services assurés par les organismes communautaires et les soins informels qui sont donnés au sein de la famille;
- l'utilisation du régime fiscal pour reconnaître le travail non rémunéré qui est essentiel au bien-être des Canadiennes et des Canadiens et à l'intégrité de la société canadienne, et pour offrir une compensation à cet égard;
- l'élaboration, en ce qui a trait au milieu de travail, à la formation et aux avantages sociaux, de politiques favorables à la famille;

- l'examen des répercussions que peut avoir, sur le soin des enfants, le fait que le ménage soit le principal lieu d'activité lucrative de la famille (ex. : ménages agricoles, ménages où la maison est aussi le lieu d'affaires de travailleuses ou travailleurs autonomes).

Les universitaires et les militantes féministes soutiennent depuis longtemps qu'une compréhension plus poussée des liens entre travail rémunéré et travail non rémunéré – entre travail lucratif et travail non lucratif – est essentielle à la compréhension de l'inégalité entre les femmes et les hommes et à l'élaboration de politiques plus équitables. Au Canada, où la capacité de mesure et d'évaluation est parmi les meilleures au monde, l'attention s'est déplacée vers les incidences politiques. La présence de questions sur le travail non rémunéré dans le recensement de 1996 ainsi que la collecte et l'analyse subséquentes des données apporteront une contribution majeure aux discussions :

- en attirant l'attention sur l'existence et l'importance du travail qui se fait à l'extérieur du marché du travail rémunéré;
- en permettant la réunion systématique d'une information de grande qualité partout au pays, ce qui présente un intérêt particulier au niveau provincial et infraprovincial, puisqu'il arrive souvent que certaines zones de responsabilité ne peuvent pas recevoir de données à la suite d'enquêtes nationales, en raison soit de la petite taille des échantillons ou du fait qu'elles sont complètement exclues des enquêtes nationales (c'est le cas p. ex. des Territoires);
- en aidant à la création du plus vaste ensemble de données à ce jour sur le travail non rémunéré au pays; cet ensemble deviendra encore plus riche et plus utile à chaque nouveau recensement du Canada, la collecte répétée des mêmes renseignements permettant, avec les années, de disposer d'analyses plus fiables et de qualité supérieure.

Condition féminine Canada souhaite faire avancer le débat public sur les questions d'égalité entre les sexes et permettre aux particuliers, aux groupes, aux décisionnaires et aux analystes de politiques de participer plus efficacement à l'élaboration des politiques. C'est avec plaisir que nous appuyons ce travail, et nous remercions les analystes de Statistique Canada de leur apport considérable et à propos.

REMERCIEMENTS

L'auteur et l'auteure aimeraient exprimer leur reconnaissance à Michel Côté et à son équipe pour leur généreuse assistance relativement à l'extraction de totalisations spéciales des fichiers principaux du recensement de 1996 ainsi qu'à Chris Jackson qui a corrigé les erreurs détectées dans les versions antérieures du rapport. L'auteur et l'auteure tiennent également à remercier tout spécialement Ian Macredie et Benoit Laroche, qui ont accepté d'affecter des ressources humaines considérables à la présente étude. Plusieurs des graphiques et des tableaux définitifs sont attribuables à l'excellent travail de Sharron Smith.



IMPORTANCE DES DONNÉES DU RECENSEMENT SUR LE TRAVAIL NON RÉMUNÉRÉ ET PRINCIPALES QUESTIONS ABORDÉES DANS L'ÉTUDE D'ÉVALUATION

Introduction

Tout ensemble de statistiques a des limites importantes. Les nouvelles formes de statistiques, en particulier, font l'objet d'améliorations au fur et à mesure qu'elles sont utilisées. Les nouvelles données du recensement de 1996 sur certains aspects du travail non rémunéré n'échappent pas à la règle. En dépit des limites, toutefois, il est possible, en faisant appel à des applications ingénieuses et originales des méthodes d'estimation (qui n'ont cependant rien de plus original que l'usage communément admis, aujourd'hui, des simulations et des données synthétiques pour appuyer l'analyse et l'élaboration de politiques), de se servir des données du recensement de 1996 sur le travail non rémunéré afin de produire des résultats de recherche applicables à un large éventail de secteurs politiques.

Précisément parce qu'il s'agit de toutes nouvelles données dans le contexte d'un recensement, il est essentiel, pour en faire une analyse fondamentale, de procéder à des études d'évaluation qui examinent les principaux aspects de la qualité de ces données. Il faut tout d'abord élaborer une notion utile de ce qu'on entend par « qualité globale », et l'appliquer à une évaluation de la qualité globale des données du recensement sur le travail non rémunéré par rapport au niveau de qualité qui caractérise les données du recensement pour des questions considérées comme ordinaires, dont l'utilisation est bien établie dans le cadre d'un recensement national. Si la qualité globale correspond au niveau ordinaire, il est avantageux de cerner les principaux sous-groupes de la population pour lesquels la qualité est soit particulièrement faible soit « supérieure à la moyenne » du point de vue de l'utilité des données. Il est aussi important de déterminer les grandes catégories d'applications analytiques qui sont bloquées en raison de particularités dans les questions du recensement ou dans la représentation subséquente des résultats dans le fichier principal du recensement de Statistique Canada. Le défi que représente l'exécution de ces évaluations a été en partie relevé dans le cadre de trois études achevées récemment, lesquelles constituent les chapitres deux, trois et quatre du présent document. L'information connexe, qui est axée sur les possibilités et les limites à des fins d'analyse, se trouve dans Stone et Silver (1998a).

Une fois qu'on a pris la décision d'utiliser le recensement pour fournir des données sur le travail non rémunéré (comme cela a été le cas avec le recensement de 1996), et qu'on a mené à bien la collecte des données, on a enclenché un nouveau processus d'évaluation qui peut avoir une portée nationale.

Principales questions abordées

Un sommaire des questions abordées dans chacun des articles est présenté ci-dessous.

Le premier article traite de la qualité des estimations des trois types de travail non rémunéré couverts par le recensement de 1996. Il s'appuie sur un examen des taux de non-réponse et de l'ampleur de l'imputation attribuable à des réponses manifestement erronées. L'article cherche à déterminer si la qualité des estimations de l'importance du travail non rémunéré axées sur les données du recensement atteint des normes minimales de qualité qui justifieraient leur utilisation dans le cadre d'analyses ultérieures visant diverses sous-populations de la société canadienne, la qualité étant évaluée uniquement à la lumière des perspectives précitées.

Le deuxième article évalue la qualité des résultats des questions du recensement de 1996 qui portent sur les travaux ménagers non rémunérés, les soins non rémunérés aux enfants et les soins non rémunérés aux personnes âgées à partir de comparaisons entre ces résultats et ceux obtenus à l'aide de questions identiques posées dans le cadre de l'Enquête sociale générale (ESG) de 1996. La question clé abordée dans cet article est la suivante : quelles sont les conséquences, sur le plan de la qualité, des écarts relevés entre les résultats de l'ESG de 1996 et ceux du recensement de 1996 lorsqu'on calcule des distributions similaires (p. ex. la distribution des répondantes et des répondants selon le nombre d'heures consacrées aux soins des enfants, par catégorie) à partir des deux sources de données, compte tenu que l'on a posé des questions identiques?

Le troisième article évalue la qualité des résultats des questions du recensement de 1996 qui portent sur les travaux ménagers non rémunérés et les soins non rémunérés aux enfants à partir de comparaisons entre ces résultats et les données de journaux personnels sur l'emploi du temps tirées de l'ESG de 1992. Cette étude cherche à déterminer si certaines tendances d'écart entre les données de journaux et les données provenant de questions directes (celles du recensement sont classées dans cette dernière catégorie), relevées dans la documentation, sont observées dans des données raisonnablement comparables de certaines distributions des résultats de l'ESG de 1992 et du recensement de 1996. Seules les questions du recensement traitant des travaux ménagers et des soins aux enfants sont abordées dans cet article, les données de journaux de l'ESG ne permettant pas d'analyser les heures consacrées aux soins donnés aux personnes âgées.

Résultats choisis

Le sommaire suivant des principaux résultats de l'ensemble des articles pourrait se révéler utile pour les lectrices et les lecteurs qui n'ont pas le temps de lire en détail les trois études.

Les deux premières études d'évaluation indiquent que les données du recensement sur le travail non rémunéré atteignent un « niveau de qualité » comparable à celui de variables du recensement nettement plus établies. Le cas le plus défavorable est celui des soins aux enfants : cette question se situe à l'extrême inférieure de la fourchette de qualité comparativement à la même question posée dans le cadre de l'ESG de 1996 (voir l'article 2). (Le terme « même question » ne se rapporte pas aux données sur les soins aux enfants tirées de journaux de l'emploi du temps mentionnées précédemment.)

Toutefois, dans le cas d'un segment important de la population (soit les adultes âgés de 25 à 64 ans ayant fait des études postsecondaires, et particulièrement celles et ceux qui sont actifs sur le marché du travail), les données du recensement et celles de l'ESG sont très comparables, une fois prises en considération les sources de non-comparabilité entre les deux ensembles de données. Pour cette population, il est difficile de déterminer quelle source de données est la meilleure. Dans la mesure où certains concepts subtils entrent dans la formulation des questions de recensement sur l'emploi du temps relatives aux soins à autrui, on peut supposer qu'un niveau de scolarité plus élevé implique des probabilités plus élevées de comprendre ces questions et d'y répondre correctement.

Les données du recensement visant certains groupes clés (les personnes sans emploi, celles qui sont peu scolarisées et les membres de groupes ethniques ou linguistiques exotiques) doivent être utilisées avec beaucoup de circonspection. Par ailleurs, pour ces mêmes groupes, la taille de l'échantillon de l'ESG est souvent tellement réduite qu'aucune des deux sources ne semble satisfaisante sur le plan de la qualité des données.

L'une des principales faiblesses des données du recensement tient à la nature des catégories de réponse se rapportant aux questions sur le travail non rémunéré. Une des tâches fondamentales consistera à modifier ces catégories pour accroître leur utilité à des fins d'analyse.

Enfin, on ne peut passer sous silence un autre des avantages du recensement : la taille considérable de l'échantillon et, de ce fait, sa capacité de produire des données infraprovinciales ou encore des données visant diverses sous-populations. Ces possibilités sont fort importantes dans la mesure où bon nombre de politiques sociales sont de portée infranationale, c'est-à-dire d'ordre provincial, infraprovincial ou même local.

Principales observations relatives aux résultats

Les résultats relevés dans les trois articles soulèvent un certain nombre d'observations générales importantes et de conclusions qui méritent d'être soulignées.

Premièrement, dans la mesure où le recensement porte uniquement sur les travaux ménagers et les soins aux enfants et aux personnes âgées, on ne peut établir aucune estimation du *volume total du travail non rémunéré* à partir de cette source de données. Par conséquent, on ne peut ni confirmer ni infirmer, à partir des données du recensement, les hypothèses selon lesquelles le groupe A, par exemple, effectuerait plus de travail non rémunéré que le groupe B, après avoir pris en considération la taille de ces groupes et d'autres facteurs dont il convient de tenir compte. Il s'agit plutôt d'utiliser ces données pour analyser divers types particuliers de travail non rémunéré. Heureusement, le recensement de 1996 comporte des questions visant des enjeux de grand intérêt.

Deuxièmement, compte tenu que les répondantes et les répondants au recensement ont fourni des renseignements en fonction de grandes catégories d'heures consacrées à divers types de travail non rémunéré et que le recensement comprend une importante catégorie ouverte, on

ne peut établir facilement des estimations des heures consacrées à des types particuliers de travail à partir des totalisations du recensement. On doit plutôt élaborer et appliquer des modèles statistiques de façon à transformer les réponses initialement données selon les grandes catégories en estimations qu'il est possible d'ajouter pour produire des totaux simulés. En ce qui a trait aux données du recensement de 1996 sur les heures consacrées aux soins aux enfants, un tel modèle a déjà été élaboré, mis à l'essai et utilisé (voir Stone et Silver, 1998b)¹.

Troisièmement, il serait possible de faire une utilisation nettement plus importante des données sur les heures consacrées aux soins des personnes âgées si l'on concevait une question supplémentaire permettant aux analystes de distinguer les adultes qui ont peu ou qui n'ont pas de possibilités d'apporter leur aide à des personnes âgées. L'absence d'une telle question restreint l'analyse des données aux cas où les répondantes et les répondants au recensement vivent avec des personnes âgées, particulièrement s'il s'agit de membres de la famille ou d'autres parentes ou parents. Heureusement, l'échantillon du recensement pour de tels ménages est considérablement plus important et plus utile aux administrations provinciales (les principales autorités responsables des politiques, des programmes et des services visant les personnes âgées) que l'échantillon de toute autre enquête de Statistique Canada.

Compte tenu que le recensement fournit également des renseignements sur l'état d'incapacité de l'ensemble des membres du ménage, cette source de données s'impose comme un élément central de l'information qui sert de fondement aux débats (et à la planification des programmes connexes) liés à l'ampleur et à la distribution des activités *non officielles* de soins à la maison chez un vaste éventail de sous-populations canadiennes.

Quatrièmement, en ce qui a trait aux soins aux enfants, il serait fort trompeur de soutenir que les estimations des heures consacrées aux soins aux enfants *à titre d'activité principale* axées sur les données de journaux de l'emploi du temps de l'ESG sont comparables (et *a fortiori* supérieures) aux données issues de la question directe du recensement (« Combien d'heures avez-vous consacrées à... »). L'article 3 démontre de façon concluante qu'il s'agit de variables fondamentalement différentes. (Un exemple analogue serait de confondre le revenu total défini comme le revenu de travail touché et le revenu total défini comme le revenu de travail et bien d'autres sources de revenu, par exemple le revenu de placement. Ces variables sont tout simplement des totaux fondamentalement différents.)

La supériorité de la question du recensement tient au fait qu'elle fait référence à un concept beaucoup plus large des soins aux enfants que ne le font les données de journaux de l'emploi du temps de l'ESG visant à mesurer uniquement les activités principales. Ce concept est nettement plus susceptible de présenter de l'intérêt pour les analystes des soins aux enfants et les mères que le concept de l'activité principale sur lequel sont axées les données de l'emploi du temps. Par conséquent, si l'ESG devait s'imposer comme une solution de recharge sérieuse au recensement, à titre de source de données sur les heures consacrées aux soins aux enfants, il faudrait utiliser une question de type censitaire dans le cadre de l'ESG (et non les journaux de l'emploi du temps de l'ESG à moins qu'ils ne soient accompagnés d'un journal sur les soins aux enfants). Il s'agit là d'une conclusion importante puisque la principale

opposition aux données du recensement sur le travail non rémunéré, dont la qualité est jugée insuffisante, est fondée sur la supériorité présumée des données tirées des journaux de l'emploi du temps. Nous soutenons que, au contraire, pour ce qui est des soins aux enfants, les données tirées de journaux de l'emploi du temps qui mesurent uniquement les activités principales sont inférieures aux données issues de la question de type censitaire.

Par conséquent, en ce qui a trait à l'accès à des données sur l'emploi du temps axées sur le concept plus large et plus utile des soins aux enfants dans un contexte où une diversité de sous-groupes clés de la population doivent être représentés par des sous-échantillons de taille suffisante, les enquêtes nationales ne constituent pas une solution de rechange au recensement. Le recensement demeure la seule source capable de fournir les renseignements requis.

Enfin, on estime généralement que l'ESG est un instrument capable fournir des données sur l'emploi du temps supérieures à celles qu'il est possible d'obtenir dans le cadre d'un recensement; toutefois, les articles démontrent qu'il est extrêmement difficile d'établir, dans le cas d'une sous-population très importante de Canadiens et de Canadiens, la supériorité présumée des données de l'ESG en ce qui a trait aux trois questions couvertes par le recensement (pour citer le cas le plus favorable à l'ESG).

ARTICLE 1 : NON-RÉPONSE ET IMPUTATION DANS LES CAS DE RÉPONSES MULTIPLES

Introduction

Le présent article présente une évaluation générale de la qualité des estimations des trois types de travail non rémunéré couverts par le recensement de 1996. Il s'appuie sur un examen des taux de non-réponse et de l'imputation des données. L'article cherche à déterminer si les estimations de l'importance du travail non rémunéré axées sur les données du recensement atteignent des normes minimales de qualité qui justifieraient leur utilisation dans le cadre d'analyses ultérieures visant diverses sous-populations de la société canadienne. (Un sommaire des résultats est présenté à la fin de l'article.)

Contexte

Les questions sur le travail non rémunéré ont été intégrées au recensement pour la première fois en 1996. Ces renseignements ont été recueillis auprès d'un échantillon de 20 % des ménages à l'aide du questionnaire (complet) 2B². Les questions sur le travail non rémunéré ont été posées à l'ensemble des répondantes et des répondants âgés de 15 ans et plus, exception faite des pensionnaires à temps plein d'établissement.

Le questionnaire comportait trois questions à cet égard. Les répondantes et les répondants étaient priés de déclarer le nombre d'heures consacrées au cours de la semaine précédant le recensement aux (1) *travaux ménagers non rémunérés*, (2) *aux soins non rémunérés aux enfants* et (3) *aux soins ou à l'aide non rémunérés aux personnes âgées*. Les répondantes et les répondants devaient cocher la case correspondant à la catégorie du nombre d'heures consacrées à chacune de ces activités. (Voir l'annexe 1.A pour un exemplaire des questions apparaissant sur le questionnaire et pour les instructions connexes présentées dans le guide.)

Un concept dont il importe de tenir compte ici est celui des « *activités simultanées* ». Ce terme désigne des périodes au cours desquelles les répondantes ou les répondants se livraient simultanément à des activités d'au moins deux des trois catégories précitées. Par exemple, des travaux ménagers accomplis en même temps que la garde d'enfants constituent des activités simultanées. Dans ces cas, les répondantes et les répondants devaient inscrire le même nombre d'heures dans les deux catégories applicables, soit sous la rubrique relative aux travaux ménagers et sous celle touchant les soins aux enfants.

Qualité des données

Une fois que l'on recevait les questionnaires remplis, les renseignements étaient saisis dans l'un des sept centres de traitement de Revenu Canada au pays. Les données saisies étaient ensuite transmises à Statistique Canada, qui les chargeait dans une base de données en vue d'un traitement ultérieur.

Avant l'étape du traitement, on examine les données d'entrée pour vérifier la compréhension des répondantes et des répondants et la difficulté de répondre aux questions de même que la réticence à répondre à des questions potentiellement délicates ou inopportunes. Les deux indicateurs servant à évaluer ces problèmes sont le taux de non-réponse et le taux de réponses multiples à des questions qui commandent une réponse unique.

Taux de non-réponse

Le taux de non-réponse correspond au nombre de personnes d'une catégorie donnée de la population qui n'ont pas répondu à une question particulière du recensement *à laquelle elles auraient dû répondre*, divisé par la population totale de cette catégorie qui aurait dû répondre à la question. Par exemple, puisque l'ensemble des répondantes et des répondants âgés de 15 ans et plus (exception faite des pensionnaires d'établissement) devaient répondre aux questions se rapportant au travail non rémunéré, le taux de non-réponse pour l'une des questions relatives au travail non rémunéré (p. ex. les travaux ménagers) est calculé ainsi :

$$\frac{\text{Nombre de personnes âgées de 15 ans et plus} \\ \text{n'ayant pas répondu à la question sur les travaux ménagers non rémunérés}}{\text{Nombre total de personnes âgées de 15 ans et plus}}$$

On peut calculer les taux de non-réponse pour chacune des questions du recensement de même que pour chacun des sous-groupes de la population, par exemple, pour l'ensemble des résidentes et des résidents de l'Ontario ou encore pour les personnes âgées de 65 ans et plus.

Un taux de non-réponse élevé (c.-à-d. un taux généralement supérieur à celui obtenu pour d'autres questions du recensement ou pour d'autres sous-groupes de la population) peut indiquer un ou plusieurs problèmes. Il se peut que les répondantes et les répondants n'aient pas bien compris la question ou aient trouvé celle-ci trop difficile et n'aient donc pas été en mesure de fournir une réponse. Les taux de non-réponse élevés peuvent également indiquer que les répondantes et les répondants ont jugé la question inopportune ou délicate et ont donc refusé d'y répondre (p. ex. les questions sur l'union libre ou le revenu). Enfin, il arrive que les personnes ne répondent pas à une question lorsqu'elles jugent que celle-ci ne s'applique pas à leur situation.

Ainsi, la question visant l'état matrimonial est posée à l'ensemble des membres du ménage, même aux jeunes enfants. Cependant, bon nombre de répondantes et de répondants laissent cette question en blanc pour leurs enfants (l'état matrimonial ne s'applique pas aux enfants) plutôt que d'inscrire la réponse « CÉLIBATAIRE ».

Le tableau 1.1 présente une comparaison des taux de non-réponse non pondérés³ des trois variables relatives au travail non rémunéré et des taux de non-réponse se rapportant à plusieurs autres questions du recensement. Le degré de difficulté et la nature délicate des questions choisies varient pour les répondantes et les répondants.

Tableau 1.1. Comparaison des taux de non-réponse à certaines questions du recensement concernant le travail non rémunéré
(Données-échantillons (20 %) – non pondérées)

Question n°	Question	Taux de non-réponse
Q5	État matrimonial	1,9*
Q6	Union de fait	4,2*
Q13	Lieu de naissance	0,9
Q17	Origine ethnique	4,1
Q24	Cours primaire/secondaire	2,9
Q25	Années d'études universitaires	1,7
Q26	Années d'études collégiales	2,2
Q30A	Travaux ménagers non rémunérés	1,5
Q30B	Soins non rémunérés aux enfants	3,9
Q30C	Soins non rémunérés aux personnes âgées	2,6
Q31	Heures de travail rémunéré	1,5
Q34	Cherchait du travail	2,9
Q36	Date du dernier travail	2,7
Q37/38	Branche d'activité	4,6
Q45	Nombre de semaines travaillées en 1995	3,9
Q47	Revenu	5,2

Note :

* D'après les données intégrales issues du Recensement de 1996.

En ce qui a trait aux taux des questions sur le travail non rémunéré, les données indiquent que la question sur les travaux ménagers non rémunérés obtient le taux de non-réponse le plus faible (1,5 %), tandis que celle sur les soins non rémunérés aux enfants affiche le taux le plus élevé (3,9 %). Le taux de non-réponse à la question visant les soins non rémunérés aux personnes âgées s'établit à 2,6 %.

Comparativement aux autres questions du recensement, la question relative aux travaux ménagers non rémunérés obtient un taux de non-réponse relativement faible, alors que celui de la question visant les soins non rémunérés aux personnes âgées se classe à peu près en milieu de liste. En revanche, le taux relatif aux soins non rémunérés aux enfants est relativement élevé et se rapproche de celui visant l'union libre et l'origine ethnique; il reste néanmoins inférieur à celui des questions se rapportant au secteur d'activité et au revenu, questions que bon nombre de répondantes et de répondants jugent difficiles.

Le faible taux de non-réponse à la question traitant des travaux ménagers semble indiquer que les répondantes et les répondants n'ont généralement pas éprouvé de difficulté à comprendre la question ou à y répondre, et que celle-ci n'a pas été jugée trop inopportune. Compte tenu de ces résultats, il est peu probable que les répondantes et les répondants aient considéré les questions sur les soins aux enfants ou aux personnes âgées comme trop délicates. Bien que bon nombre de répondantes et de répondants aient pu éprouver de la difficulté à déterminer ce qui constitue des soins, nous posons l'hypothèse que les taux élevés de non-réponse à ces deux questions (notamment à la question relative aux soins non rémunérés aux enfants) sont sans doute davantage attribuables à la tendance qu'ont certaines personnes à ne pas répondre aux questions qui, selon elles, ne s'appliquent pas à leur situation. Par exemple, des personnes sans enfant pourraient avoir laissé en blanc la question visant les soins aux enfants plutôt que de cocher la case « AUCUNE », même si certaines d'entre elles ont pu donner des soins non rémunérés aux enfants d'autrui.

Le tableau 1.2 présente les taux de non-réponse non pondérés des variables relatives au travail non rémunéré ainsi qu'une comparaison entre les taux des femmes et des hommes dans chacune des provinces et dans chacun des territoires. Selon les données de ce tableau :

- Les hommes affichaient des taux de non-réponse supérieurs à ceux des femmes.
- Le Yukon obtenait les taux de non-réponse les plus élevés pour les trois variables.
- Les Territoires du Nord-Ouest se classaient au deuxième rang après le Yukon pour la question touchant les travaux ménagers, et la Colombie-Britannique prenait la deuxième place pour les questions se rapportant aux soins non rémunérés aux enfants et aux personnes âgées.
- Terre-Neuve, le Québec et le Nouveau-Brunswick obtenaient généralement les taux de non-réponse les plus faibles pour les trois variables.

Tableau 1.2. Taux de non-réponse, selon le sexe, pour les variables liées au travail non rémunéré – Canada, provinces et territoires
(Données-échantillons (20 %) – non pondérées)

	Travaux ménagers			Soins aux enfants			Soins aux personnes âgées		
	Total	Hommes	Femmes	Total	Hommes	Femmes	Total	Hommes	Femmes
Canada	1,5	1,6	1,4	3,9	4,1	3,6	2,6	2,7	2,4
Terre-Neuve	1,0	1,0	1,0	2,9	3,0	2,8	1,7	1,7	1,6
Île-du-Prince-Édouard	1,4	1,5	1,2	4,2	4,5	3,9	2,6	2,8	2,4
Nouvelle-Écosse	1,3	1,3	1,2	4,2	4,5	3,9	2,4	2,5	2,4
Nouveau-Brunswick	1,4	1,5	1,3	3,6	4,1	3,2	2,2	2,4	2,0
Québec	1,3	1,3	1,2	2,7	2,8	2,5	2,2	2,2	2,2
Ontario	1,5	1,7	1,4	4,3	4,6	4,0	2,7	2,9	2,5
Manitoba	1,2	1,3	1,1	3,4	3,7	3,2	2,0	2,2	1,9
Saskatchewan	1,2	1,3	1,2	3,6	3,8	3,4	2,2	2,3	2,1
Alberta	1,6	1,7	1,4	4,2	4,5	3,8	2,5	2,7	2,3
Colombie-Britannique	1,9	2,1	1,8	5,0	5,4	4,7	3,3	3,6	3,1
Yukon	5,1	5,6	4,4	6,7	7,4	5,9	5,7	6,4	4,8
Territoires du Nord-Ouest	2,2	2,3	2,1	3,2	3,4	3,1	2,7	2,8	2,6

Un examen plus détaillé de ces résultats met en relief certains éléments. Le fait que les hommes affichent des taux de non-réponse systématiquement plus élevés que les femmes peut tendre à confirmer notre hypothèse voulant que certaines personnes laissent en blanc les questions qui, selon elles, ne s'appliquent pas à leur situation. Il ne serait pas surprenant de constater que certains répondants estiment que les questions sur le travail non rémunéré s'appliquent uniquement aux femmes.

Tableau 1.3. Comparaison des taux de non-réponse à certaines questions du recensement concernant le travail non rémunéré – Canada, provinces et territoires
(Données-échantillons (20 %) – non pondérées)

	Lieu de naissance	Années d'études universitaires	Années d'études collégiales	Travaux ménagers non rémunérés	Soins non rémunérés aux enfants	Soins non rémunérés aux personnes âgées	Heures de travail rémunéré	Cherchant du travail	Date du dernier travail	Industrie	Nombre de semaines travaillées en 1995
Canada	0,9	1,7	2,2	1,5	3,9	2,6	1,5	2,9	2,7	4,6	3,9
Terre-Neuve	0,3	0,9	1,3	1,0	2,9	1,7	1,1	1,8	1,4	4,4	3,4
Île-du-Prince-Édouard	0,6	1,6	2,2	1,4	4,2	2,6	1,6	3,0	2,6	4,4	3,6
Nouvelle-Écosse	0,5	1,4	1,9	1,3	4,2	2,4	1,3	2,4	2,0	4,6	3,5
Nouveau-Brunswick	0,9	1,5	1,9	1,4	3,6	2,2	1,4	2,6	2,2	4,7	3,6
Québec	0,7	1,4	2,0	1,3	2,7	2,2	1,3	2,7	2,4	3,9	3,7
Ontario	1,0	1,8	2,3	1,5	4,3	2,7	1,6	3,1	2,8	5,1	4,0
Manitoba	0,6	1,4	1,8	1,2	3,4	2,0	1,3	2,6	2,2	3,7	3,2
Saskatchewan	0,6	1,4	1,8	1,2	3,6	2,2	1,4	2,9	2,5	4,4	3,4
Alberta	0,8	1,7	2,1	1,6	4,2	2,5	1,6	3,0	2,7	4,4	3,8
Colombie-Britannique	1,2	2,2	2,7	1,9	5,0	3,3	2,0	3,6	3,3	5,2	4,7
Yukon	3,7	5,2	5,6	5,1	6,7	5,7	4,8	7,0	6,1	6,7	6,4
Territoires du Nord-Ouest	1,5	2,6	3,1	2,2	3,2	2,7	2,2	3,9	3,5	3,2	3,5

Cependant, la possibilité que les hommes n'établissent pas de liens entre les travaux ménagers qu'ils accomplissent et les questions sur le travail non rémunéré a des répercussions sur la qualité des estimations du recensement : le nombre d'heures que les hommes consacrent au travail non rémunéré est-il sous-estimé? La distribution des données du recensement pour les hommes se révélera particulièrement intéressante lorsque nous examinerons la qualité des estimations axées sur le recensement dans les rapports subséquents.

Deuxièmement, le taux élevé de non-réponse observé au Yukon n'est pas une situation propre aux questions sur le travail non rémunéré. Comme l'indique le tableau 1.3, dans le cas de toutes les questions du recensement pour lesquelles nous disposons de données provinciales, le Yukon obtient les taux de non-réponse les plus élevés. De façon analogue, les Territoires du Nord-Ouest et la Colombie-Britannique se disputaient en alternance les deuxième et troisième places à ce chapitre. Par conséquent, s'il faut faire preuve de circonspection lorsqu'on utilise les données sur le travail non rémunéré au Yukon, cette mise en garde doit également s'appliquer aux autres variables du recensement.

Taux de réponses multiples

Le second indicateur servant à évaluer la compréhension des répondantes et des répondants et leur réaction à une question particulière du recensement est le taux de réponses multiples à une question à laquelle une réponse unique est exigée. Les réponses multiples erronées se produisent lorsqu'une personne donne plus d'une réponse valide à une question qui commande une réponse unique. (Par exemple, dans le cas du travail non rémunéré, la répondante ou le répondant coche plus d'une catégorie d'« heures » pour une question donnée relative au travail non rémunéré.) Ce concept permet de définir ainsi le « taux de réponses multiples erronées » : le nombre de réponses multiples erronées pour 10 000 réponses fournies.

Tableau 1.4. Comparaison des taux de réponse multiple à certaines questions du recensement concernant le travail non rémunéré
(Données-échantillons (20 %) – non pondérées)

Question n°	Question	Taux de réponse multiple (par tranche de 10 000 réponses)
Q5	État matrimonial	10*
Q13	Lieu de naissance	28
Q30A	Travaux ménagers non rémunérés	77
Q30B	Soins non rémunérés aux enfants	39
Q30C	Soins non rémunérés aux personnes âgées	18
Q34	Cherchait du travail	31
Q36	Date du dernier travail	63
Q41	Catégorie de travailleur	51

Note :
* D'après les données intégrales issues du Recensement de 1996.

Le tableau 1.4 présente le taux de réponses multiples non pondéré se rapportant aux questions du recensement sur le travail non rémunéré et le compare au taux d'autres questions du recensement de forme similaire (c.-à-d. des questions auxquelles les personnes doivent répondre en cochant une des réponses suggérées).

En fait, et c'est là un élément clé, le niveau absolu de réponses multiples erronées est extrêmement faible, soit à peine suffisant pour avoir une incidence sur toute statistique fondée sur un échantillon d'au moins 1 000 répondantes et répondants. À titre d'exemple, la question sur les travaux ménagers non rémunérés obtenait le taux le plus élevé (soit environ 77 réponses multiples erronées pour 10 000 réponses) tandis que les questions portant sur les soins aux enfants et aux personnes âgées affichaient environ 39 et 18 réponses multiples erronées pour 10 000 réponses, respectivement.

Bien que les taux de réponses multiples erronées soient très faibles en nombres absolus, la question se rapportant aux travaux ménagers non rémunérés obtient le taux le plus élevé de l'ensemble des variables du recensement pour lesquelles des données existent. (La question relative à la date du dernier emploi se classe au deuxième rang à ce chapitre avec un taux de 63 réponses multiples erronées pour 10 000 réponses.) En revanche, les taux des deux autres questions sur le travail non rémunéré sont parmi les plus faibles figurant au tableau 1.4. Les taux les plus faibles se rapportent à l'état matrimonial et au lieu de naissance, deux questions pour lesquelles on s'attend à des niveaux extrêmement faibles de réponses multiples erronées. Ces résultats indiquent que les données relatives à au moins deux des questions du recensement portant sur le travail non rémunéré sont de qualité acceptable, si l'on tient compte de cet indicateur particulier.

Rien ne permet d'affirmer ici que la qualité des données sur les travaux ménagers non rémunérés est insatisfaisante, dans la mesure où le niveau absolu de cet indicateur reste extrêmement faible. Cependant, les résultats obtenus tendent à indiquer que cette variable mérite une attention particulière, ce dont il faudra tenir compte au moment d'évaluer la qualité des données au moyen d'autres indicateurs.

Le taux de réponses multiples plus élevé qu'obtient la question sur les travaux ménagers pourrait indiquer que les répondantes et les répondants éprouvent certaines difficultés à calculer le nombre d'heures consacrées à cette activité. Cette question englobe une gamme plus vaste d'activités que les questions sur les soins aux enfants et les soins aux personnes âgées, de sorte que les répondantes et les répondants peuvent se montrer indécis quant aux activités à inclure ou avoir de la difficulté à se souvenir du nombre d'heures consacrées à ces travaux. Par ailleurs, si les répondantes et les répondants avaient, dans un premier temps, inclus les heures consacrées aux soins aux enfants ou aux personnes âgées dans la catégorie des travaux ménagers, il se peut que ces personnes aient corrigé par la suite leur réponse initiale. Si une telle correction n'a pas été observée au moment de la saisie des données du questionnaire, on obtient alors une réponse multiple.

Vérification et imputation

Le terme « vérification et imputation » s'entend du processus de correction des erreurs de non-réponse et de réponses multiples dans le cadre du recensement.

La vérification consiste à examiner les réponses saisies en vue de détecter des erreurs, telles la non-réponse et les réponses multiples erronées. Comme on l'a mentionné précédemment,

ces erreurs peuvent être attribuables au fait que les répondantes et les répondants ont donné des réponses erronées ou incomplètes aux questions ou il peut s'agir d'erreurs d'entrée commises au moment de la saisie des données figurant sur les questionnaires.

L'imputation est le processus qui consiste à modifier les réponses ou à inscrire de nouvelles réponses dans le but de corriger les erreurs détectées au cours de la vérification. Il existe deux grandes catégories d'imputation : l'imputation déterministe et l'imputation stochastique. Dans le cas de l'imputation déterministe, on corrige les erreurs en déduisant la valeur appropriée à partir des réponses données à d'autres questions du recensement. Par exemple, si la question portant sur le sexe est restée sans réponse, mais que la personne a déclaré ailleurs être le fils de la personne 1, on peut alors imputer la valeur « masculin », soit la réponse correcte pour la variable correspondant au sexe de la personne.

Pour expliquer l'imputation stochastique, prenons l'exemple d'une donnée manquante ou manifestement erronée, qu'il n'est pas possible d'imputer au moyen de l'imputation déterministe. Supposons que les réponses saisies d'une personne indiquent que celle-ci travaille à l'Hôpital d'Ottawa, mais que la question sur la profession soit restée sans réponse. En principe, l'imputation stochastique consiste à chercher au hasard, dans les enregistrements du recensement déjà corrigés, une autre personne considérée comme « suffisamment similaire en ce qui a trait à des aspects pertinents » et qui a donné une réponse à la question relative à la profession. La réponse de cette personne est alors inscrite sur l'enregistrement sous la rubrique de la profession pour laquelle les données étaient manquantes. On désigne par le terme « donneur » la personne qui fournit l'information. Le terme « imputation par enregistrement donneur » utilisé ci-dessous désigne ce type d'imputation.

Il convient d'élucider la formulation « suffisamment similaire en ce qui a trait à des aspects pertinents ». Pour simplifier la chose, poursuivons avec notre exemple. Puisque nous savons déjà que la personne travaille à l'Hôpital d'Ottawa, nous pourrions décider que le fait de travailler dans un hôpital constitue un aspect pertinent et, par conséquent, seules les personnes ayant un tel emploi seront considérées comme des donneurs potentiels. En général, un ensemble de caractéristiques, ou « critères d'appariement », que l'on juge connexes aux données manquantes sont sélectionnées, et l'on procède ensuite à la recherche de personnes qui correspondent à la répondante ou au répondant dont l'enregistrement comporte une erreur quant à ces caractéristiques. Ces personnes sont considérées comme des donneurs. Les renseignements manquants sont alors imputés à partir des enregistrements donneurs.

Quelquefois, on n'arrive pas à établir un appariement complet en fonction de tous les critères. Dans ces cas, on détermine le meilleur appariement possible compte tenu des paramètres établis de la recherche⁴.

Les observations précédentes visaient à initier les lectrices et les lecteurs aux principaux aspects de l'imputation. Dans les paragraphes qui suivent, nous expliquons comment l'imputation a été appliquée aux types d'erreur présentés plus tôt.

Imputation dans les cas de non-réponse

Dans le cas des répondantes et des répondants âgés de 15 ans et plus qui n'ont pas répondu à une ou plusieurs questions sur le travail non rémunéré, on a attribué des valeurs au moyen de l'imputation stochastique. L'imputation des variables du travail non rémunéré s'est faite en deux volets. On a utilisé des processus distincts pour les questions visant les travaux ménagers non rémunérés et les soins non rémunérés aux enfants, d'une part, et les soins non rémunérés aux personnes âgées, d'autre part. On a estimé que, pour le choix d'un enregistrement donneur, les caractéristiques susceptibles de déterminer le nombre d'heures consacrées aux travaux ménagers ou aux soins aux enfants (p. ex. la présence d'enfants au sein du ménage) seraient moins pertinentes pour évaluer le nombre d'heures consacrées aux soins aux personnes âgées. Le fait de séparer le processus d'imputation en deux volets a permis d'utiliser des critères d'appariement différents pour les deux ensembles de variables.

Afin de trouver un enregistrement donneur permettant d'imputer les données manquantes sur les travaux ménagers et les soins aux enfants, on a d'abord stratifié les enregistrements selon le type de logement. Les personnes vivant dans des logements collectifs ont, en outre, été stratifiées selon le type de logement collectif. Les personnes vivant dans des logements privés ont été stratifiées selon le sexe et le groupe d'âge. Dans les diverses strates définies par le type de logement, le sexe et l'âge, on a retenu les variables suivantes comme critères supplémentaires d'appariement pour sélectionner les enregistrements donneurs : groupes d'âge définis par tranches de cinq ans, état matrimonial, présence d'enfants, nombre d'enfants, nombre d'heures de travail rémunéré et lieu de travail.

Le processus d'imputation de la variable visant les soins aux personnes âgées est fondé sur une stratification des personnes vivant dans des logements collectifs selon le type de logement collectif. Les personnes vivant dans des logements privés, toutefois, ont été stratifiées selon le sexe et l'appartenance à une famille économique comptant une personne âgée (soit une personne âgée de 60 ans et plus). Les personnes des familles économiques ne comptant pas de personnes âgées ont, en outre, été classées en deux catégories d'âge : les personnes âgées de moins de 45 ans et les personnes âgées de 45 ans et plus. Les critères d'appariement utilisés pour choisir les donneurs sont les suivants : groupes d'âge définis par tranches de cinq ans, état matrimonial, lieu de travail, nombre d'heures de travail rémunéré et groupe d'âge de la personne âgée (dans le cas des répondantes et des répondants vivant avec une personne âgée).

Au terme des processus d'imputation, un enregistrement donneur caractérisé par une concordance complète, en fonction de tous les critères d'appariement, avec l'enregistrement comportant une erreur a été trouvé dans environ 92 % des cas⁵ pour les variables des travaux ménagers et des soins aux enfants et dans 98 % des cas pour la variable des soins aux personnes âgées. Dans les deux processus d'imputation, on a réussi à trouver des donneurs pour l'ensemble des enregistrements.

Le tableau 1.5 présente la distribution des variables relatives au travail non rémunéré avant et après le processus d'imputation stochastique. Les données, qui visent l'ensemble du Canada, indiquent des changements minimes des distributions résultant des procédures d'imputation

employées. Il importe de noter que, pour les sous-populations, les changements des distributions provoqués par le processus d'imputation restent peu marqués.

Tableau 1.5. Comparaison des variables liées au travail non rémunéré avant et après imputation de la non-réponse, Canada
(Données-échantillons (20 %) – non pondérées)

	Avant imputation	Après imputation de la non-réponse	Taux de répartition avant et après imputation
Travaux ménagers non rémunérés :			
Total	100,0	100,0	
Aucune	11,9	12,0	1,01
Moins de 5 heures	22,2	22,3	1,00
De 5 à 14 heures	29,9	29,9	1,00
De 15 à 29 heures	19,2	19,1	0,99
De 30 à 59 heures	11,6	11,6	1,00
60 heures et plus	5,2	5,2	1,00
Soins non rémunérés aux enfants :			
Total	100,0	100,0	
Aucune	60,4	60,9	1,01
Moins de 5 heures	9,7	9,6	0,99
De 5 à 14 heures	10,3	10,2	0,99
De 15 à 29 heures	7,0	6,9	0,99
De 30 à 59 heures	5,4	5,4	1,00
60 heures et plus	7,2	7,1	0,99
Soins non rémunérés aux personnes âgées :			
Total	100,0	100,0	
Aucune	83,0	83,0	1,00
Moins de 5 heures	10,9	10,8	0,99
De 5 à 9 heures	3,4	3,4	1,00
10 heures et plus	2,7	2,7	1,00

Résolution des réponses multiples erronées

Les erreurs prenant la forme de réponses multiples aux questions sur le travail non rémunéré ont été corrigées au moyen de l'imputation déterministe. Dans tous les cas, la catégorie comptant le moins d'heures indiquée par les répondantes et les répondants a été retenue. On a opté pour cette méthode afin d'éviter l'affectation d'un nombre disproportionné de personnes à la catégorie supérieure du « nombre d'heures » et pour conserver l'une des réponses initiales inscrites sur le questionnaire.

Le tableau 1.6 présente la distribution des données avant et après le processus d'imputation déterministe. Puisque le nombre de réponses multiples erronées est extrêmement faible comparativement au nombre total de réponses, l'imputation a eu une incidence minime sur la distribution des données.

Conclusion

Le rapport évalue, dans un premier temps, la qualité des estimations du travail non rémunéré établies à partir des données du recensement de 1996. Il s'appuie sur un examen des taux de non-réponse et de réponses multiples et de l'imputation subséquente des données. Les résultats obtenus indiquent que, comparativement à d'autres questions du recensement, la question sur les travaux ménagers non rémunérés donne lieu à un taux relativement faible de non-réponse. Cependant, les questions sur les soins non rémunérés aux personnes âgées, tout particulièrement, et sur les soins non rémunérés aux enfants, obtiennent des taux de non-réponse relativement plus élevés. Nous posons l'hypothèse que les taux plus élevés de non-réponse à ces deux questions sont vraisemblablement attribuables à la tendance qu'ont

certaines personnes à ne pas répondre aux questions qui, selon elles, ne s'appliquent pas à leur situation.

Les hommes affichaient des taux de non-réponse systématiquement plus élevés que les femmes dans l'ensemble des provinces et des territoires. Si ces résultats peuvent tendre à confirmer l'hypothèse précitée, ils soulèvent la possibilité que les hommes soient moins susceptibles d'établir des liens entre les travaux ménagers qu'ils accomplissent et les questions du recensement sur le travail non rémunéré et qu'ils sous-estiment, en définitive, le nombre d'heures consacrées à ces activités.

Tableau 1.6. Comparaison des variables liées au travail non rémunéré
avant et après imputation des réponses multiples
(Données-échantillons (20 %) – non pondérées)

	Avant imputation	Après imputation de la non-réponse	Taux de répartition avant et après imputation
Travaux ménagers non rémunérés :			
Total	100,0	100,0	
Aucune	11,7	11,9	1,02
Moins de 5 heures	22,2	22,2	1,00
De 5 à 14 heures	30,0	29,9	1,00
De 15 à 29 heures	19,2	19,2	1,00
De 30 à 59 heures	11,6	11,6	1,00
60 heures et plus	5,3	5,2	0,98
Soins non rémunérés aux enfants :			
Total	100,0	100,0	
Aucune	60,5	60,4	1,00
Moins de 5 heures	9,7	9,7	1,00
De 5 à 14 heures	10,3	10,3	1,00
De 15 à 29 heures	6,9	7,0	1,01
De 30 à 59 heures	5,4	5,4	1,00
60 heures et plus	7,2	7,2	1,00
Soins non rémunérés aux personnes âgées :			
Total	100,0	100,0	
Aucune	83,0	83,0	1,00
Moins de 5 heures	10,9	10,9	1,00
De 5 à 9 heures	3,4	3,4	1,00
10 heures et plus	2,7	2,7	1,00

L'évaluation des taux de réponses multiples erronées indique que les données associées aux variables des soins aux enfants et des soins aux personnes âgées sont de qualité acceptable. En revanche, la question se rapportant aux travaux ménagers non rémunérés obtenait le taux le plus élevé de l'ensemble des variables du recensement pour lesquelles des données étaient disponibles. Toutefois, dans la mesure où le nombre absolu de réponses multiples demeure extrêmement faible, on ne doit pas considérer que les données sur les travaux ménagers non rémunérés sont de qualité inacceptable. Ces résultats peuvent plutôt indiquer que cette variable mériterait qu'on s'y attarde dans le cadre d'analyses ultérieures.

On a fait appel à une imputation déterministe consistant à retenir la catégorie inférieure du « nombre d'heures » cochée par les répondantes et les répondants pour résoudre les cas de

réponses multiples erronées. Une imputation stochastique ou par enregistrement donneur a été utilisée pour corriger la non-réponse. Ce processus d'imputation a une incidence minime sur la distribution des données qui en résulte, tant à l'échelle du Canada que pour les sous-groupes provinciaux de la population.

Annexe 1.a : questions du recensement sur le travail non rémunéré et instructions du guide

ACTIVITÉS À LA MAISON		43.
<p>Nota : Par « semaine dernière », on entend la période de sept jours (du dimanche au samedi) précédent le jour du recensement.</p> <p>À la question 30, dans les cas où les activités se déroulent en même temps, inscrivez la durée commune qui leur est consacrée dans plus d'une section.</p> <p>30. La semaine dernière, combien d'heures cette personne a-t-elle consacrées aux activités suivantes :</p> <p>a) aux travaux ménagers, à l'entretien de la maison ou du jardin pour les membres du ménage ou pour d'autres personnes, sans rémunération?</p> <p>Par exemple, préparer les repas, faire la lessive, faire la comptabilité, effectuer des courses ou tondre le gazon.</p>		
	01	— Aucune
	02	— Moins de 5 heures
	03	— De 5 à 14 heures
	04	— De 15 à 29 heures
	05	— De 30 à 59 heures
	06	— 60 heures et plus
<p>b) à s'occuper de ses enfants ou des enfants des autres, sans rémunération?</p> <p>Par exemple, donner le bain à de jeunes enfants ou jouer avec eux, conduire les enfants à leurs activités sportives, les aider à faire leurs devoirs, parler avec les adolescents de leurs problèmes.</p>	07	— Aucune
	08	— Moins de 5 heures
	09	— De 5 à 14 heures
	10	— De 15 à 29 heures
	11	— De 30 à 59 heures
	12	— 60 heures et plus

c) à offrir des soins ou de l'aide à une ou plusieurs personnes âgées, sans rémunération?	13	—	Aucune
Par exemple, donner des soins personnels à une personne âgée faisant partie de la famille, visiter des personnes âgées, leur parler au téléphone, les aider à effectuer leurs emplettes, leurs opérations bancaires ou à prendre leurs médicaments.	14	—	Moins de 5 heures
	15	—	De 5 à 9 heures
	16	—	10 heures et plus

Instructions du guide

ACTIVITÉS À LA MAISON

La question 30 fournira des renseignements sur le temps que les personnes de 15 ans et plus consacrent aux travaux ménagers, à prendre soin des enfants et à dispenser des soins et de l'aide aux personnes âgées. Ces renseignements permettront de mieux comprendre de quelle façon ces activités non rémunérées viennent contribuer au bien-être de la population canadienne.

QUESTION 30 – Travail non rémunéré

Cochez le cercle correspondant au nombre total d'heures consacrées la semaine dernière, sans recevoir de rémunération, à chacune des activités indiquées dans les parties a) à c).

Comptez les heures consacrées à ces activités non rémunérées pour :

- des membres du ménage;
- d'autres membres de la famille ne faisant pas partie du ménage;
- des amis ou des voisins.

Ne comptez pas les heures consacrées à ces activités :

- pour une rémunération (déclarez le travail rémunéré à la question 31);
- comme bénévole non rémunéré pour un organisme à but non lucratif, un organisme religieux, une œuvre de charité ou un groupe communautaire.

Activités se déroulant en même temps

Les gens font souvent plus d'une activité non rémunérée à la fois. Par exemple, il se peut qu'une personne passe une heure à préparer un repas tout en s'occupant de son ou ses enfants. En pareil cas, cette personne doit déclarer une heure de travaux ménagers à la partie a) et une heure de soins aux enfants à la partie b) de la question 30 et faire abstraction du fait que les deux activités se sont déroulées simultanément.

Partie a) – Le travail non rémunéré pour l'exécution de tâches ménagères et l'entretien du terrain et du logement

Aucunes instructions additionnelles.

Partie b) – Le soin non rémunéré des enfants

Comptez les heures passées à parler ou à jouer avec les enfants si, au cours de celles-ci, la personne visée était responsable du bien-être des enfants en question.

Partie c) – Le travail non rémunéré pour offrir des soins aux personnes âgées

Les personnes âgées sont toutes les personnes de 65 ans et plus ainsi que certaines personnes ayant près de 65 ans et souffrant d'incapacités associées au vieillissement.

ARTICLE 2 : QUALITÉ DES DONNÉES DU RECENSEMENT DE 1996 SUR LE TRAVAIL NON RÉMUNÉRÉ À LA LUMIÈRE DE COMPARAISONS AVEC LES DONNÉES DE L'ESG DE 1996 FONDÉES SUR DES QUESTIONS IDENTIQUES

Introduction

Le présent article évalue la qualité des résultats des questions du recensement de 1996 qui portent sur les travaux ménagers non rémunérés, les soins non rémunérés aux enfants et les soins non rémunérés aux personnes âgées à partir de comparaisons entre ces résultats et ceux obtenus à l'aide de questions identiques posées dans le cadre de l'Enquête sociale générale (ESG) de 1996. La question clé abordée dans cet article est la suivante : quelles sont les conséquences, sur le plan de la qualité, des écarts relevés entre les résultats de l'ESG de 1996 et ceux du recensement de 1996 lorsqu'on calcule des distributions similaires (p. ex. la distribution des répondantes et des répondants selon le nombre d'heures consacrées aux soins des enfants, par catégorie) à partir des deux sources de données, compte tenu que l'on a posé des questions identiques?

L'article est le deuxième d'une série de rapports traitant de l'évaluation de la qualité et du « potentiel d'utilisation » des questions du recensement sur le travail non rémunéré. Le premier article de la série porte sur la qualité des données du recensement déterminée en fonction des taux de non-réponse et des niveaux d'imputation attribuables à des réponses manifestement erronées (Swain et Stone, 1998). Le troisième rapport présente les résultats d'une évaluation de la qualité fondée sur l'utilisation des données tirées des journaux de l'emploi du temps de l'ESG de 1992.

Méthodologie

L'indice de dissimilitude (ID) est la statistique clé ayant servi à établir les comparaisons présentées ci-dessous. On doit calculer une distribution en pourcentage (pour une variable donnée) à partir des résultats de l'ESG, répéter ces calculs visant la même variable à partir des données du recensement, puis mesurer l'écart global entre ces deux distributions. L'ID correspond à la moitié de la somme des écarts absolus entre les distributions apparemment comparables issues des deux enquêtes. Il s'agit d'un indice bien connu des statisticiennes et des statisticiens.

Il convient, dans un premier temps, de déterminer la valeur de l'ID pour l'ensemble des échantillons du recensement et de l'ESG (qui représentent la population âgée de 15 ans et plus). Une fois cette valeur définie, on doit déterminer si elle indique des lacunes graves, sur le plan de la qualité, dans une ou dans les deux sources de données.

Il est possible de poser certaines questions statistiques types et d'y répondre lorsqu'on cherche à déterminer si une valeur donnée de l'ID indique une divergence « significative » entre les deux sources de données. On pourrait effectuer un test de signification statistique, fondé à tout le moins sur des techniques « bootstrap » qui neutralisent la distribution de

l'ESG. Ce test permettrait de déterminer si la valeur de l'ID est aisément attribuable au hasard, dans le cas où la distribution de l'ESG serait la distribution correcte. Nous évitons d'utiliser un tel test parce que, dans la plupart des cas, la taille de l'échantillon, du moins pour ce qui est du recensement, est tellement importante que des valeurs de l'ID peu élevées pourraient sembler significatives. Même si seule la signification statistique était mise en question, bon nombre d'utilisatrices et d'utilisateurs de ces résultats pourraient subtilement gonfler l'importance de ces données jusqu'à leur accorder une signification réelle.

Par conséquent, nous utilisons deux autres méthodes pour établir le cadre qui nous permettra de déterminer si une valeur donnée de l'ID tend à indiquer un écart substantiel entre les résultats du recensement et ceux de l'ESG. La première méthode soulève la question suivante. Supposons que la distribution de l'ESG est la distribution correcte. Quel niveau d'erreur de classification des réponses recueillies dans le cadre du recensement nous signalerait un danger sérieux que les données du recensement soient trompeuses?

La réponse à cette question est arbitraire (comme l'est d'ailleurs la définition d'un seuil de signification dans un test de signification en bonne et due forme). Cependant, les lignes directrices relatives aux communiqués officiels des données établissent à 15 % ou moins le coefficient de variation qui correspond à une variabilité d'échantillonnage suffisamment faible pour diffuser des données jugées raisonnablement fiables. Si nous utilisons cette règle, nous estimerons que l'une des deux sources de données est vraisemblablement gravement altérée par des erreurs de réponse uniquement lorsque l'ID sera supérieur à 15 %.

Bien entendu, il convient de considérer qu'un ID de 15 % reste élevé du point de vue de l'utilité des données. Toutefois, comme nous l'expliquerons plus loin, il ne faut pas sauter à la conclusion que le problème réside entièrement dans les données du recensement.

La seconde méthode employée pour élaborer un cadre d'interprétation d'une valeur de l'ID unique consiste à calculer la valeur de l'ID pour les variables démographiques les plus courantes (p. ex. l'âge ou l'état matrimonial). Le choix de variables qui sont généralement mesurées de façon assez précise nous permet de définir l'étalement en fonction duquel nous déterminerons qu'une valeur donnée de l'ID pour les résultats relatifs au travail non rémunéré signifie probablement des problèmes graves sur le plan de la qualité des données de l'une des deux sources.

L'étude des variations chez divers sous-groupes de la population est une activité importante axée sur les données du recensement. Par conséquent, le présent article fait état des résultats obtenus par le calcul de la valeur de l'ID pour plusieurs sous-populations. Cependant, le but principal de l'examen des variations visant ces sous-groupes est de mettre en relief les populations pour lesquelles les données du recensement semblent particulièrement faibles et les populations pour lesquelles ces données se comparent avantageusement à celles de l'ESG. Ainsi, dans le cadre de la discussion des variations des sous-groupes, nous retenons des populations particulières qui se distinguent quant à la qualité apparente, médiocre ou bonne, des données du recensement s'y rapportant. Nous ne cherchons pas à examiner la tendance de variation des valeurs de l'ID pour diverses distributions de la population dans son ensemble. La

discussion est donc très restreinte comparativement à la portée des renseignements figurant dans les tableaux suivants qui présentent les valeurs de l'ID.

Les données sont présentées en trois grandes catégories, une pour chacune des questions du recensement sur le travail non rémunéré. Avant d'analyser les données, toutefois, il convient de faire certaines mises en garde quant à la supériorité présumée des données de l'ESG. Ces mises en garde se rapportent aux principales sources de non-comparabilité relative à la méthodologie et à la période de référence du recensement de 1996 et de l'ESG de 1996.

Comparabilité des données du recensement et de l'ESG

L'ESG de 1996 est la seule enquête comportant les trois questions sur le travail non rémunéré telles que celles-ci apparaissent dans le recensement de 1996. Elle offre donc une possibilité nette de comparer et d'évaluer la qualité des résultats du recensement. Parmi ces deux enquêtes, on considère généralement que l'ESG fournit des données supérieures, et ce, pour plusieurs raisons présentées ci-dessous.

On doit faire preuve de prudence lorsqu'on interprète les écarts entre les résultats de l'ESG et ceux du recensement comme une indication de l'infériorité des données du recensement. Bien que le recensement et l'ESG comportent des questions identiques permettant de recueillir des données sur le travail non rémunéré, plusieurs différences entre les deux enquêtes peuvent avoir une incidence directe sur les résultats obtenus, laquelle peut remettre en question la comparabilité entre les deux sources.

La première différence importante se rapporte à la méthode de collecte des données. L'ESG est réalisée par interviews téléphoniques, tandis que le recensement est réalisé par autodénombrement. On sait fort bien que cette seule différence peut influer sur les données. En effet, les répondantes et les répondants peuvent modifier leur réponse à une question posée au téléphone si on leur accorde quelques minutes pour réfléchir et donner une réponse correcte. Par exemple, si on pose la question « Quel était votre revenu au cours de la dernière année civile? », les répondantes et les répondants pourraient donner une réponse quelconque si on leur accorde 15 secondes et une réponse complètement différente si ces personnes prennent le temps de consulter leurs dossiers d'impôt.

On estime généralement que les réponses fournies dans le cadre de l'ESG sont plus fiables parce que les questions sont posées par des intervieweuses et des intervieweurs formés qui sont immédiatement à la disposition des répondantes et des répondants éprouvant des difficultés à comprendre les questions sur le travail non rémunéré ou à y répondre. Les intervieweuses et les intervieweurs peuvent conseiller les personnes, leur expliquer quelles activités doivent être considérées sous chacune des rubriques du travail non rémunéré et les aider à calculer les heures consacrées à ces activités.

Par ailleurs, bien que les répondantes et les répondants au recensement remplissent le questionnaire du recensement sans l'aide d'une intervieweuse ou d'un intervieweur, la méthode de l'autodénombrement leur permet de prendre le temps de réfléchir aux réponses et de consulter d'autres membres du ménage. Compte tenu qu'il ne s'agit pas d'une interview

téléphonique, les personnes se sentent moins pressées de donner une réponse rapidement. Les répondantes et les répondants qui éprouvent des difficultés à répondre aux questions sur le travail non rémunéré peuvent consulter un guide présentant des renseignements supplémentaires. (On croit généralement que ce guide est rarement consulté.) Un système d'assistance téléphonique est également mis à la disposition des répondantes et des répondants qui désirent parler directement à des recenseuses ou recenseurs. Ces personnes offrent le même type d'assistance que les intervieweuses et les intervieweurs. Toutefois, il est possible que les répondantes et les répondants se montrent moins enclins à faire appel à ce type d'aide qu'aux conseils des intervieweuses et des intervieweurs.

La période de référence à laquelle se rapportent les activités non rémunérées déclarées constitue une deuxième différence importante entre le recensement et l'ESG. L'ESG recueille des données sur une période d'un an, alors que le recensement recueille des données à une date précise. L'ESG a été réalisée mensuellement de février à décembre 1996. Les répondantes et les répondants ont déclaré leur travail non rémunéré pour la semaine précédant l'interview. Les questions du recensement sur le travail non rémunéré ne visent que les activités s'étant déroulées au cours de la semaine précédant le jour du recensement, soit le 14 mai.

Par conséquent, les données de l'ESG, mais non celles du recensement, tiennent compte des variations saisonnières relatives à la diversité des activités déclarées de même qu'au nombre d'heures que les personnes tendent à consacrer à certaines activités. Si les données de l'ESG offrent certes un avantage du fait qu'elles représentent la situation pour toute l'année, cette caractéristique constitue, cependant, un désavantage dans le cas de la présente étude. En effet, les données devraient, idéalement, refléter la situation au cours d'une période de référence restreinte et commune à l'ensemble des répondantes et des répondants. À la limite, la comparabilité des estimations axées sur le recensement et sur l'ESG est compromise.

Une autre raison invoquée pour justifier la supériorité des données de l'ESG par rapport à celles du recensement, sur le plan de l'exactitude, tient au fait que l'ESG ne permet pas les déclarations par procuration dans le cas des trois questions touchant le travail non rémunéré. L'ESG recueille des données uniquement auprès d'une personne par ménage. La personne choisie déclarait les heures qu'elle avait consacrées, elle-même, à chacun des types de travail non rémunéré. Si la personne choisie n'était pas disponible pour passer l'interview, on ne posait pas les questions relatives au travail non rémunéré. Dans le cadre du recensement, par contre, on présume qu'une personne membre du ménage répond aux questions sur le travail non rémunéré (et à toutes les autres questions du recensement) au nom de l'ensemble des membres du ménage (déclaration par procuration).

L'incidence relative de la déclaration par procuration ne peut être évaluée pour ces sources de données dans la mesure où l'ESG ne cherche pas à recueillir des renseignements sur l'emploi du temps auprès de l'ensemble des membres du ménage des répondantes et des répondants. Pour établir une comparaison entre ces deux sources, il aurait fallu que les deux enquêtes recueillent des données visant l'ensemble des membres des ménages chantonnés. Tant que l'ESG ne recueillera pas de données pour l'ensemble des membres des ménages de

l'échantillon, on ne pourra pas déterminer ce que serait un niveau comparable de déclaration par procuration dans le cadre de l'ESG.

Par ailleurs, dans le recensement, la répondante ou le répondant qui remplit le questionnaire a le temps de consulter les autres membres du ménage au besoin. En outre, le questionnaire du recensement peut être rempli par plus d'une personne, ce qui ne peut se produire dans le cas de l'ESG tant que l'on utilisera la méthode de l'interview téléphonique pour recueillir les données.

En fait, nous ne disposons d'aucune estimation de l'ampleur des déclarations par procuration faites dans le cadre du recensement pour lesquelles la personne qui a rempli le questionnaire aurait omis d'effectuer les vérifications appropriées auprès des autres membres du ménage avant de répondre en leur nom.

La taille de l'échantillon est l'un des aspects de l'ESG qui ne peut être jugé supérieur au recensement. Compte tenu de la taille considérable de son échantillon, le recensement permet de produire des estimations plus fiables que celles de l'ESG sur le plan de la variabilité de l'échantillonnage pour des douzaines de sous-populations importantes. L'ESG s'est appuyée sur un échantillon d'environ 12 000 personnes âgées de 15 ans et plus vivant dans les dix provinces, exception faite des pensionnaires d'établissement. L'échantillon du recensement était composé d'un ménage sur cinq au Canada. (Les questions du recensement sur le travail non rémunéré étaient posées aux membres du ménage âgés de 15 ans et plus, exception faite des pensionnaires d'établissement.) Par conséquent, l'échantillon du recensement est nettement plus important que celui de l'ESG, de sorte qu'il permet de produire des estimations caractérisées par une faible variation d'échantillonnage pour des douzaines de sous-populations, ce que l'ESG n'est pas en mesure de faire.

Enfin, les différences considérables de traitement de la non-réponse dans les deux enquêtes sont telles que la comparabilité des estimations du recensement et de l'ESG pourrait être entachée. Dans le cadre du recensement, on impute des données pour la non-réponse à toutes les questions, tandis que l'ESG n'impute des données que dans le cas des variables démographiques clés comme l'âge, le sexe et la situation des particuliers dans le ménage. Par conséquent, presque toutes les variables de la base de données de l'ESG comportent un élément de non-réponse. Pour certaines sous-populations, le niveau de non-réponse est suffisamment élevé pour remettre en question l'exactitude des estimations de l'ESG. On devrait, à tout le moins, tenir compte du taux de non-réponse de l'ESG lorsqu'on observe des variances considérables entre les données du recensement et de l'ESG. Lorsque le taux de non-réponse de l'ESG correspond à un pourcentage élevé de l'écart global entre le recensement et l'ESG, nous devons nous montrer sceptiques quant à l'hypothèse soutenant la supériorité des données de l'ESG.

Bref, nous estimons que l'ESG produit vraisemblablement des données plus exactes que le recensement pour des estimations générales à l'échelle du pays. Cependant, l'importance de cette supériorité quant à l'exactitude des données ne peut être établie compte tenu des différences précitées sur le plan de la méthodologie d'enquête.

Lorsque nous examinons plusieurs sous-populations, les sources de non-comparabilité entre les deux ensembles de données sont tellement considérables qu'il serait imprudent de conclure à la supériorité des données de l'ESG, d'autant plus que les variations observées dans les sous-populations sont d'une importance capitale dans de nombreux domaines des sciences sociales. Lorsque de telles variations constituent une dimension fondamentale de l'analyse, il convient de faire preuve de scepticisme quant à la supériorité présumée des données de l'ESG.

Par ailleurs, on ne devrait pas croire que la valeur des données du recensement sur le travail non rémunéré réside d'abord dans la production d'estimations générales à l'échelle du pays si l'ESG a vraisemblablement l'avantage à ce chapitre. On peut soutenir que, compte tenu de ce que l'on sait déjà au sujet de ces agrégations à la lumière des données de l'ESG, de nouvelles estimations visant de grands regroupements de population seraient nettement moins utiles et profitables que de nouvelles données visant des sous-populations pour lesquelles l'ESG est incapable de produire des estimations utilisables.

Écarts globaux

Dans le cas des trois variables relatives au travail non rémunéré, les valeurs globales de l'ID sont inférieures au seuil de 15 %, ce qui indiquerait des divergences graves entre le recensement et l'ESG (tableau 2.1). L'écart le plus important est observé dans les données sur les soins aux enfants (ID de 11 %). Lorsqu'on compare ce résultat aux valeurs de l'ID calculées pour un « ensemble étalon » de variables démographiques (tableau 2.1), seul le plus haut niveau de scolarité atteint affiche un ID plus élevé, de près de 13 %. Le revenu du ménage est l'autre variable caractérisée par un ID relativement important (11 %). Avec un ID de 9 %, les travaux ménagers non rémunérés suivent la variable du revenu du ménage quant à l'ampleur des divergences entre les données du recensement et celles de l'ESG, tandis que l'ID des soins non rémunérés aux personnes âgées (2 %) se classe parmi les valeurs les plus faibles et s'apparente à l'ID de l'état matrimonial.

Par conséquent, si l'on tient compte de ces comparaisons globales, il semble que, pour ce qui est des variables du recensement relatives au travail non rémunéré, on obtienne un niveau d'exactitude se situant dans une fourchette comparable à celle d'autres variables mesurées régulièrement dans le cadre du recensement.

Comme on l'a noté précédemment, les valeurs de l'ID relatives aux travaux ménagers non rémunérés et aux soins non rémunérés aux personnes âgées s'établissent à 9 % et à 2 %, respectivement. Chez les femmes, les chiffres correspondants sont de 11 % et de 3 %, respectivement. Ces valeurs de l'ID sont plus élevées que les valeurs correspondantes chez les hommes. Il s'agit d'un résultat qui va à l'encontre de l'hypothèse voulant que les données du recensement visant les hommes soient particulièrement faibles comparativement aux données analogues de l'ESG.

Tableau 2.1. Indices globaux de dissimilitude entre la répartition d'un ensemble comparatif de variables selon le recensement et selon l'ESG

Indice de dissimilitude pour le recensement et l'ESG	Recensement	ESG	Différence
Travaux ménagers non rémunérés :			
Aucune	11,5	7,1	4,4
Moins de 5 heures	22,7	22,9	0,2
De 5 à 14 heures	30,3	37,8	7,5
De 15 à 29 heures	19,2	21,1	1,9
De 30 à 59 heures	11,4	9,4	2
60 heures et plus	4,8	1,6	3,1
Non précisé*		4	
Indice de dissimilitude			9,5
Soins non rémunérés aux enfants (ménages comptant (au moins un enfant de moins de 15 ans) :			
Aucune	18,7	13,2	5,5
Moins de 5 heures	12,7	9,4	3,3
De 5 à 14 heures	20,6	18,1	2,5
De 15 à 29 heures	16,5	20	3,5
De 30 à 59 heures	13,5	19,3	5,8
60 heures et plus	18	20	2
Non précisé*		3,3	
Indice de dissimilitude			11,2
Soins non rémunérés aux personnes âgées :			
Aucune	83,6	81,1	2,4
Moins de 5 heures	10,8	11,5	0,7
De 5 à 9 heures	3,2	3,7	0,4
10 heures et plus	2,4	3,7	1,3
Non précisé*		2,8	
Indice de dissimilitude			2,4
Groupes d'âge :			
De 15 à 24 ans	17	17	0
De 25 à 44 ans	41,2	41,7	0,5
De 45 à 64 ans	27,3	26,9	0,4
65 ans et plus	14,5	14,5	0
Indice de dissimilitude			0,5
État matrimonial (estimations du recensement fondées sur la totalité des données) :			
Jamais marié(e)	27	25,6	1,4
Marié(e)/conjoint(e) de fait	59,1	61,6	2,5
Séparé(e)/divorcé(e)	7,7	6,8	0,9
Veuf(ve)	6,2	6	0,2
Non précisé*		0,4	
Indice de dissimilitude			2,5

Tableau 2.1 (suite)

Indice de dissimilitude pour le recensement et l'ESG	Recensement	ESG	Différence
Lieu de naissance :			
Dans la province de résidence	64,4	65,8	1,3
Dans une autre province	14,2	14,4	0,2
États-Unis	1	1,3	0,3
Amérique centrale et Amérique du Sud	2,3	2	0,3
Royaume-Uni	2,9	3,3	0,4
Autres pays d'Europe	7,3	6,9	0,4
Afrique	1	0,9	0,1
Asie	6,6	5,2	1,4
Occéanie et autres pays	0,2	0,2	0
Non précisé*		3,3	
Indice de dissimilitude			2,2
Langue parlée à la maison :			
Anglais	66	64,9	1,1
Français	22,7	23,4	0,7
Autre langue	9,3	5,4	3,9
Anglais et français	0,4	1,7	1,3
Anglais et autre langue	1,4	3,8	2,4
Anglais, français et autre langue**	0,2	0,9	0,7
Non précisé*		3,4	
Indice de dissimilitude			5,1
Plus haut niveau de scolarité atteint :			
Universitaire	13,3	15,2	1,9
Diplôme d'études collégiales	16,4	9,7	6,7
Certificat ou diplôme d'une école de métiers	10,5	13,7	3,2
Etudes postsecondaires partielles	10,8	15,8	5
Certificat d'études secondaires	14,3	16,8	2,5
Etudes primaires/études secondaires partielles	34,8	28,9	5,9
Non précisé*		3,7	
Indice de dissimilitude			12,6
Profession : (1)			
A. Gestion	9,4	10,8	1,4
B. Commerce, finance et administration	19,4	17	2,3
C. Sciences naturelles et appliquées	5,1	5,3	0,2
D. Santé	5,3	6,4	1,1
E. Sciences sociales, enseignement, services gouvernementaux et services connexes	7	7	0
F. Arts, culture, loisirs et sports	2,7	2,9	0,2
G. Vente et services	25,8	25,5	0,3
H. Métiers, transport et opérateurs de machines	13,5	13,4	0,1
I. Industries primaires	4,5	3,7	0,8
J. Transformation, fabrication et services publics	7,4	7,9	0,6
Non précisé*		3,6	
Indice de dissimilitude			3,5

Tableau 2.1 (suite)

Revenu personnel total :			
Aucun revenu	7,6	8	0,4
Moins de 5 000 \$	13,9	10,1	3,9
5 000 \$ à 9 999 \$	11,8	10,7	1
10 000 \$ à 14 999 \$	12,9	11,4	1,5
15 000 \$ à 19 999 \$	9,2	7,8	1,4
20 000 \$ à 29 999 \$	15	17,1	2,2
30 000 \$ à 39 999 \$	11,4	12,6	1,2
40 000 \$ à 49 999 \$	7,4	8,8	1,5
50 000 \$ à 59 999 \$	4,6	5,9	1,2
60 000 \$ à 79 999 \$	3,8	4,6	0,8
80 000 \$ à 99 999 \$	1,1	1,3	0,2
100 000 \$ et plus	1,3	1,6	0,4
Ne sait pas/non précisé*		26,3	
Indice de dissimilitude			7,8
Revenu du ménage :			
Aucun revenu ou revenu perdu	0,3	0,5	0,2
Moins de 10 000 \$	5,7	3,8	1,9
10 000 \$ à 19 999 \$	11,1	13,1	2
20 000 \$ à 29 999 \$	11,8	14,7	2,9
30 000 \$ à 39 999 \$	11,8	14,2	2,4
40 000 \$ à 49 999 \$	11,5	14	2,4
50 000 \$ à 59 999 \$	10,7	11,7	1
60 000 \$ à 79 999 \$	16,5	13	3,4
80 000 \$ à 99 999 \$	9,6	7,6	2
100 000 \$ et plus	11,1	7,4	3,6
Ne sait pas/non précisé*		33,8	
Indice de dissimilitude			11

Notes :

*Les valeurs non précisées figurent à titre informatif seulement.

Elles ne sont pas comprises dans l'estimation de la répartition.

**Comprend les catégories « français et autre langue » et « anglais, français et autre langue ».

(1) Univers du recensement = population active occupée, y compris les absents.

Univers de l'ESG = personnes qui travaillaient pendant la semaine précédant l'enquête – exclut les absents.

La valeur exceptionnellement faible de l'ID des soins non rémunérés aux personnes âgées est partiellement attribuable au nombre peu élevé de catégories de réponse possibles. Par ailleurs, ces données pourraient se révéler plus utiles, à titre d'indicateur des divergences entre le recensement et l'ESG, si elles ne visaient que les personnes d'âge mûr pour lesquelles les soins aux personnes âgées sont susceptibles de constituer un problème de la vie quotidienne. Cependant, comme l'indique le tableau 2.2, on observe peu de différences dans les résultats lorsque les données sont restreintes aux répondantes et aux répondants âgés de 45 à 64 ans.

Pour ce qui est de la direction apparente du biais dans les distributions des données du recensement, on observe, pour les trois variables relatives au travail non rémunéré, que les données du recensement tendent systématiquement à afficher une forte pondération à l'extrême

inférieure de la distribution (soit la catégorie du nombre d'heures « aucune »). À l'extrême supérieure, on observe une « surpondération » apparente des distributions du recensement dans le cas des travaux ménagers et, chez les hommes seulement, dans le cas des soins aux enfants.

Tableau 2.2. Nombre d'heures consacrées aux soins non rémunérés aux personnes âgées par les personnes de 15 ans et plus, selon l'âge et le sexe

	Total			25 à 44 ans		
	Recensement	ESG	Différence	Hommes		
				Recensement	ESG	Difference
Aucune	86,4	84,0	2,4	87,0	84,0	3,0
Moins de 5 heures	9,6	10,1	0,5	9,4	10,6	1,2
De 5 à 9 heures	2,4	3,4	1,0	2,2	3,2	1,0
10 heures et plus	1,6	2,5	0,9	1,4	2,1	0,7
Non précisé		3,0			3,0	
Indice de dissimilitude			2,4			3,0
	45 à 64 ans			65 ans et plus		
Aucune	82,5	82,0	0,6	86,4	82,1	4,4
Moins de 5 heures	12,1	10,2	1,9	7,8	10,7	2,9
De 5 à 9 heures	3,3	4,6	1,4	2,9	3,4	0,5
10 heures et plus	2,1	3,2	1,1	2,9	3,9	1,0
Non précisé		3,1			4,4	
Indice de dissimilitude			2,4			4,4
	Total			25 à 44 ans		
	Recensement	ESG	Différence	Recensement	ESG	Difference
				Femmes		
Aucune	80,8	78,4	2,4	80,6	79,6	1,0
Moins de 5 heures	12,0	12,9	0,9	12,9	12,2	0,7
De 5 à 9 heures	4,1	4,0	0,1	3,9	3,8	0,1
10 heures et plus	3,1	4,7	1,6	2,6	4,4	1,8
Non précisé		2,7			2,1	
Indice de dissimilitude			2,5			1,8
	45 à 64 ans			65 ans et plus		
Aucune	75,0	72,1	2,9	83,4	79,1	4,2
Moins de 5 heures	13,4	16,3	1,9	8,8	11,1	2,4
De 5 à 9 heures	5,9	5,0	0,9	4,0	4,5	0,5
10 heures et plus	4,7	6,6	1,9	3,9	5,2	1,3
Non précisé		3,0			4,7	
Indice de dissimilitude			3,8			4,2

Note :

En raison de la faible taille des cellules de l'ESG, le tableau ne montre pas les données relatives aux répondants de 15 à 24 ans.

Sous-groupes de la population caractérisés par des niveaux anormalement élevés de divergence entre le recensement et l'ESG

Travaux ménagers

Les groupes d'âge inférieur et supérieur (personnes âgées de 15 à 24 ans et de 65 ans et plus, respectivement) obtiennent des ID nettement plus élevés que la moyenne. Il convient à cet égard de souligner les taux élevés de non-réponse dans l'ESG pour le groupe des personnes âgées de 65 ans et plus : 7 % chez les hommes et 10 % chez les femmes. De plus, les personnes vivant seules ou qui n'avaient pas travaillé au cours de la semaine précédant l'enquête ou qui ne parlaient pas l'anglais à la maison affichaient également des ID supérieurs à la moyenne.

Le tableau 2.3 indique que, chez les hommes, l'ID n'est supérieur à la moyenne que dans le cas des hommes vivant seuls. Cet indice est inférieur à la moyenne pour les hommes et les femmes qui vivent en couple et qui ont une enfant ou un enfant à la maison. Chez les femmes, l'ID est supérieur à la moyenne pour celles qui vivent seules et pour les mères seules qui vivent avec une enfant ou un enfant âgé de moins de 15 ans. Ces résultats semblent aller à l'encontre de l'hypothèse selon laquelle un niveau plus élevé de déclarations par procuration dans le cadre du recensement tend à réduire la qualité des données de celui-ci par rapport aux données de l'ESG.

Dans le cas des hommes comme des femmes qui occupaient un emploi rémunéré au cours de la semaine précédant l'enquête, l'ID est nettement inférieur à la moyenne et bien moins élevé que l'ID des personnes qui n'occupaient pas un tel emploi. Ce dernier écart est tellement marqué qu'on pourrait supposer que les analyses excluant les personnes sans emploi rémunéré au cours de la semaine précédant l'enquête pourraient s'appuyer sur des données du recensement de qualité pratiquement équivalente à celle des données de l'ESG.

Cette observation est particulièrement vraie dans le cas des hommes puisque la propension de ceux-ci à occuper un emploi rémunéré est généralement très forte partout au pays. Si elle est exacte, cette proposition constituerait une autre raison nous empêchant de soutenir que la qualité des données du recensement visant les hommes est spécialement faible par rapport à la qualité des données de l'ESG.

Il convient aussi de souligner la valeur peu élevée de l'ID pour les hommes et les femmes ayant suivi des études universitaires. Ces résultats tendent à indiquer que lorsqu'il est possible de restreindre l'analyse aux personnes ayant suivi des études universitaires et qui occupaient un emploi rémunéré au cours de la semaine précédant l'enquête, la qualité des données du recensement est susceptible d'être équivalente à celle des données de l'ESG.

Tableau 2.3. Nombre d'heures consacrées aux travaux ménagers dans les ménages privés, selon le sexe et la situation des particuliers dans le ménage

Heures de travaux ménagers non rémunérés	Recensement	ESG	Difference	Recensement	ESG	Difference	Recensement	ESG	Difference	Personne avec conjoint
Aucune	15,4	10,1	5,3	15,0	7,4	7,6	13,0	11,3	1,7	
Moins de 5 heures	30,1	29,9	0,2	29,6	36,3	6,7	24,0	25,8	1,8	
De 5 à 14 heures	32,8	40,5	7,7	36,5	43,6	7,1	34,5	40,8	6,3	
De 15 à 29 heures	14,3	14,9	0,6	13,4	10,4	3,0	17,6	16,6	1,0	
30 heures et plus	7,4	4,6	2,8	5,5	2,3	3,2	10,9	5,5	5,4	
Non précisée		3,9			5,0			3,8		
Indice de dissimilarité			8,3			13,8			8,1	
<i>Personne avec conjoint et au moins un enfant est âgé de 15 ans et plus</i>										
										Hommes
Aucune	7,9	6,4	1,5	1,3,4	16,4	3,0				
Moins de 5 heures	24,6	21,2	3,4	26,3	21,4	4,9				
De 5 à 14 heures	39,4	45,6	6,2	35,9	38,2	2,3				
De 15 à 29 heures	19,1	20,0	0,9	16,2	17,9	1,7				
30 heures et plus	9,1	6,8	2,3	8,2	6,2	2,0				
Non précisée		4,3			3,1			6,9		
Indice de dissimilarité				7,1						

Tableau 2.3 (suite)

Heures de travaux ménagers non rémunérés	Recensement		ESG		Difference		Recensement	ESG	Difference
	Total	Personne avec conjoint	ESG	Difference	Personne avec conjoint et au moins un enfant de moins de 15 ans				
Personne avec conjoint, dont le plus jeune enfant est âgé de 15 ans et plus									
Aucune	7,7	4,3	3,4	5,4	4,7	0,7	1,8	2,4	0,6
Moins de 5 heures	15,7	16,1	0,4	9,9	12,0	2,1	4,7	5,7	1,0
De 5 à 14 heures	28,1	35,3	7,2	31,1	39,5	8,4	22,5	27,3	4,8
De 15 à 29 heures	24,0	27,0	3,0	27,6	29,8	2,2	30,4	34,2	3,8
De 30 à 59 heures	17,0	14,4	2,6	20,3	11,4	8,9	24,9	24,8	0,1
60 heures et plus	7,6	2,8	4,8	5,7	2,7	3,0	15,6	5,7	9,9
Non précisé		4,2			5,3			3,1	
Indice de dissimilarité		10,7			12,6			10,1	
Personne seule									
Aucune	12,9	6,6	6,3	3,6	3,7	0,1	7,0	3,4	3,6
Moins de 5 heures	19,8	21,0	1,2	6,3	10,0	3,7	10,5	14,1	3,6
De 5 à 14 heures	35,5	45,4	9,9	25,5	21,8	3,7	30,5	36,1	5,6
De 15 à 29 heures	19,8	20,1	0,3	30,9	37,5	6,6	27,3	28,3	1,0
30 heures et plus	12,0	6,9	5,1	33,7	27,0	6,7	24,7	18,1	6,6
Non précisé		7,4			3,6			4,4	
Indice de dissimilarité			11,4		10,4				
Parent seul ayant au moins un enfant de moins de 15 ans									
Aucune	2,6		1,1	1,5					
Moins de 5 heures	6,3		8,9	2,6					
De 5 à 14 heures	26,2		34,2	8,0					
De 15 à 29 heures	29,2		32,2	3,0					
De 30 à 59 heures	21,4		18,6	2,8					
60 heures et plus	14,3		5,0	9,3					
Non précisé			2,5						
Indice de dissimilarité					13,6				

Note :
En raison de la faible taille des cellules de l'ESG, les données relatives aux pères seuls n'ont pu être comparées. Les catégories « de 30 à 59 heures » et « 60 heures et plus » ont été regroupées dans le cas des femmes seules, des femmes vivant avec un conjoint et un enfant de 15 ans et plus, et des parents seuls vivant avec un enfant de 15 ans et plus.

Soins aux enfants

On observe, pour les soins aux enfants, des tendances généralement similaires à celles des travaux ménagers en ce qui a trait à l'ID. Une différence marquée, toutefois, se rapporte aux mères seules, la divergence entre le recensement et l'ESG étant inférieure à la moyenne dans le cas des soins aux enfants.

Les données sur les soins aux enfants se limitent aux personnes vivant avec une enfant ou un enfant âgé de moins de 15 ans, et il convient de porter une attention particulière au groupe des 25 à 44 ans (tableau 2.4). Dans ce groupe d'âge, l'ID pour les hommes s'établit à 15 %, soit deux points au-dessus de la moyenne. L'ID pour les femmes atteint 13 %, soit la moyenne. Pour l'un et l'autre sexe, la valeur de l'ID est largement attribuable à la « surpondération » apparente des distributions du recensement dans la catégorie inférieure du nombre d'heures consacrées aux soins aux enfants.

Ces résultats tendent à indiquer que les données du recensement ne surestiment pas les activités de soins aux enfants dispensés par les hommes ou les femmes. L'écart plus marqué observé chez les hommes semble peu lié aux déclarations par procuration (si l'on présume que la conjointe remplit le questionnaire) puisque les deux sexes affichaient une « surpondération » prononcée des distributions du recensement dans la catégorie inférieure de l'échelle.

Pour les deux sexes, les personnes vivant en couple obtiennent un ID équivalent à la moyenne, soit 13 %, de sorte que la valeur de 15 % citée précédemment pour les hommes et supérieure à la moyenne est attribuable aux hommes dont la situation dans le ménage est différente.

En résumé, les valeurs de l'ID dans le cas des soins aux enfants se rapprochent généralement du seuil critique de 15 %. Cependant, elles ne sont pas nettement plus élevées chez les hommes que chez les femmes lorsqu'on considère uniquement les personnes vivant en couple. Les mères seules obtiennent un ID inférieur à la moyenne globale (pour les soins aux enfants). S'établissant à 10 %, l'ID relatif aux mères seules se situe nettement dans la fourchette établie de divergence entre le recensement et l'ESG qui englobe une gamme représentative de variables du recensement.

Soins aux personnes âgées

On observe, pour les soins aux personnes âgées, des tendances généralement similaires à celles des travaux ménagers présentées plus haut en ce qui concerne l'ID. Plus particulièrement, on note un ID supérieur à la moyenne dans le cas des hommes vivant uniquement avec une conjointe, des mères seules, des personnes qui n'occupaient pas d'emploi au cours de la semaine précédant l'enquête et des personnes pour lesquelles l'anglais n'est pas la langue usuelle à la maison. Toutefois, l'ID se rapportant aux soins aux personnes âgées est tellement faible (il s'établit presque invariablement sous le seuil des 5 %) (voir le tableau 2.5) qu'il ne semble pas nécessaire de discuter plus à fond cette variable. Les données indiquent que, en ce qui a trait aux soins aux personnes âgées, la qualité des données du recensement est équivalente à celle des données de l'ESG.

Tableau 2.4. Nombre d'heures consacrées aux soins non rémunérés aux enfants par les personnes de 15 ans et plus, selon le sexe et le groupe d'âge

(Ménages comptant au moins un enfant de moins de 15 ans)

	Total			25 à 44 ans			45 ans et plus		
	Recensement	ESG	Déférence	Recensement	ESG	Déférence	Recensement	ESG	Déférence
Hommes									
Aucune	23,1	16,8	6,3	12,0	6,0	6,1	25,5	22,3	3,1
Moins de 5 heures	17,8	14,0	3,8	15,9	10,8	5,0	23,5	17,5	6,0
De 5 à 14 heures	25,6	24,7	0,9	29,5	27,8	1,7	27,0	24,8	2,2
De 15 à 29 heures	16,7	25,1	8,4	21,1	30,5	9,4	12,5	22,0	9,5
De 30 à 59 heures	9,3	13,5	4,2	12,1	17,7	5,6	11,5	13,4	1,9
60 heures et plus	7,5	6,0	1,5	9,4	7,2	2,2			
Non précisée		3,8			4,3			5,1	
Indice de dissimilarité		12,6				15,0			11,4
Femmes									
Aucune	14,8	9,9	4,8	6,6	2,6	4,0	22,0	17,2	4,9
Moins de 5 heures	8,2	5,3	2,8	5,1	2,2	3,0	13,6	11,6	2,1
De 5 à 14 heures	16,1	12,2	4,0	16,2	11,2	5,1	24,1	24,6	0,6
De 15 à 29 heures	16,4	15,4	1,0	18,9	18,7	0,2	16,6	9,1	7,5
De 30 à 59 heures	17,3	24,5	7,2	20,9	26,6	5,7	23,7	37,6	13,9
60 heures et plus	27,2	32,6	5,4	32,2	38,8	6,6			
Non précisée		2,8			2,9			4,8	
Indice de dissimilarité		12,6				12,3			14,5

Note :

En raison de la faible taille des cellules de l'ESG, le tableau ne montre pas les données relatives aux répondants de 15 à 24 ans. Les données relatives aux personnes de 45 à 64 ans et à celles de 65 ans et plus ont été regroupées. Les catégories « de 30 à 59 heures » et « 60 heures et plus » ont été regroupées dans le cas des hommes et des femmes de 45 ans et plus.

Tableau 2.5. Nombre d'heures consacrées aux soins non rémunérés aux personnes âgées dans les ménages privés, selon la situation des particuliers dans le ménage

	Total			Personne seule			Personne avec conjoint		
	Recensement		ESG	Recensement		ESG	Recensement		ESG
				Hommes					
Aucune	86,4	84,0	2,4	89,3	85,9	3,4	86,6	81,1	5,5
Moins de 5 heures	9,6	10,1	0,5	7,1	8,6	1,5	9,0	11,9	2,9
De 5 à 9 heures	2,4	3,4	1,0	2,2	2,4	0,3	2,6	3,7	1,1
10 heures et plus	1,7	2,5	0,9	1,5	3,1	1,7	1,8	3,3	1,5
Non précisé		3,0			3,3			3,0	
Indice de dissimilarité			2,4			3,4			5,5
Personne avec conjoint et au moins un jeune enfant est âgé de 15 ans et plus									
	Hommes			Hommes			Hommes		
	Recensement		ESG	Recensement		ESG	Recensement		ESG
Aucune	83,4	83,4	0,0	81,9	81,6	0,2			
Moins de 5 heures	12,3	10,2	2,2	13,1	13,2	0,1			
De 5 à 9 heures	2,8	4,2	1,4	5,0	5,1	0,1			
10 heures et plus	1,5	2,3	0,8						
Non précisé		3,4			2,6				
Indice de dissimilarité			2,2			0,2			

Tableau 2.5 (suite)

	Total			Personne seule			Personne avec conjoint		
	Recensement	ESG	Difference	Recensement	ESG	Difference	Recensement	ESG	Difference
Femmes									
Aucune	80,9	78,4	2,4	83,8	79,2	4,6	80,9	78,3	2,6
Moins de 5 heures	12,0	12,9	0,9	10,0	11,4	1,3	11,1	13,0	1,8
De 5 à 9 heures	4,1	4,0	0,1	3,9	4,8	0,9	4,4	3,9	0,5
10 heures et plus	3,1	4,7	1,6	2,3	4,7	2,4	3,5	4,8	1,3
Non précisé		2,7			3,9			3,5	
Indice de dissimilarité						4,6			3,1
Personne avec conjoint, dont le plus jeune enfant est âgé de 15 ans et plus									
1 child under the age of 15									
	Total			Personne seule			Personne avec conjoint		
	Recensement	ESG	Difference	Recensement	ESG	Difference	Recensement	ESG	Difference
Femmes									
Aucune	77,8	80,7	2,9	73,6	71,1	2,6	81,0	77,7	3,4
Moins de 5 heures	14,9	12,6	2,4	16,1	16,7	0,6	11,7	9,9	1,8
De 5 à 9 heures	4,5	2,7	1,8	5,9	5,1	0,8	4,3	6,8	2,5
10 heures et plus	2,8	4,0	1,2	4,4	7,2	2,8	3,0	5,6	2,6
Non précisé		2,6			1,6			1,9	
Indice de dissimilarité		4,1			3,4			5,1	

Note :

En raison de la faible taille des cellules de l'ESG, les données relatives aux pères seuls et aux mères seules ayant un enfant de 15 ans et plus n'ont pu être comparées.

Au terme de cette discussion et tout juste avant d'aborder l'évaluation axée sur les données de l'ESG sur l'emploi du temps (on considère à ce chapitre que la supériorité de l'ESG fait peu de doute), il importe de souligner que les résultats de l'ESG sur l'emploi du temps ne comportent pas de données similaires à celles du recensement sur les soins aux personnes âgées. On ne dispose d'aucun moyen permettant d'extraire des estimations utiles des heures consacrées aux soins aux personnes âgées à partir des enquêtes sur l'emploi du temps menées en 1986 et en 1992 (ESG). Par conséquent, pour ce qui est de ce type de soins, les données du recensement sortent, sans contredit, gagnantes.

De toute évidence, les données du recensement sur les soins aux personnes âgées comportent des limites importantes dans la mesure où le recensement ne permet pas de déterminer quelles personnes avaient des parents âgés susceptibles d'avoir besoin d'assistance. Toutefois, lorsqu'on restreint l'analyse aux répondantes et aux répondants vivant avec une personne âgée, particulièrement avec une personne âgée déclarée handicapée, il est possible d'utiliser de façon avantageuse les données du recensement relatives aux soins aux personnes âgées, ce qui reste tout à fait impossible dans le cas de l'ESG. (Pour une discussion des questions connexes, voir Stone et Silver, 1998a).

Conclusion

En conclusion, seules les données relatives aux soins aux enfants se caractérisent par des ID qui se rapprochent du seuil critique indiquant la possibilité de résultats d'une qualité inacceptable. Cependant, même les données touchant les soins aux enfants obtiennent des valeurs d'ID situées dans la fourchette de valeurs observables dans l'ensemble des variables du recensement dont l'inclusion dans le recensement est jugée légitime et n'est pas remise en question.

En outre, lorsqu'il est possible de restreindre l'analyse aux personnes qui occupaient un emploi ou qui avaient suivi des études universitaires, on accroît la qualité du sous-ensemble des données du recensement utilisé, du moins comparativement à l'ensemble complet des données du recensement.

En fait, dans le cas des données sur les travaux ménagers, si l'on tient compte de cette restriction, il est impossible de déterminer quelles données d'enquête (le recensement ou l'ESG) sont de meilleure qualité en ce qui concerne les questions directes sur l'emploi du temps. Cette observation, peut-être controversée, s'applique à l'ensemble des données relatives aux soins aux personnes âgées. Cette question est tout particulièrement importante dans la mesure où les données de l'ESG tirées de journaux de l'emploi du temps (lesquelles suscitent les arguments les plus favorables à l'ESG) ne permettent en aucune façon de produire des estimations des heures consacrées aux soins aux personnes âgées.

Enfin, contrairement à ce à quoi on aurait pu s'attendre, rien dans la présente étude n'indique que les données du recensement sur le travail non rémunéré sont de moindre qualité pour les hommes que pour les femmes.

En résumé, on ne peut soutenir, en ce qui a trait à la qualité, l'argument défavorable aux données du recensement sur le travail non rémunéré à la lumière des résultats sur les tendances de non-réponse ou sur la fréquence des réponses multiples erronées (soit l'objet de notre premier article). Le présent (deuxième) article démontre qu'on ne peut soutenir un tel argument en se fondant sur la question directe sur l'emploi du temps en dépit du fait que les deux enquêtes comportent des questions identiques à cet égard. Par conséquent, l'argument défavorable aux données du recensement relatives au travail non rémunéré, pour ce qui est de la qualité, devra donc entièrement s'appuyer sur notre analyse des données tirées de journaux de l'emploi du temps recueillies dans le cadre de l'ESG de 1992. (Bien qu'il serait plus avantageux d'utiliser les données de journaux de l'emploi du temps tirées de l'ESG de 1998, celles-ci ne sont pas encore disponibles.) Cette question fera l'objet de l'article suivant.

ARTICLE 3 : QUALITÉ DES DONNÉES DU RECENSEMENT DE 1996 SUR LE TRAVAIL NON RÉMUNÉRÉ À LA LUMIÈRE DE COMPARAISONS AVEC LES DONNÉES RELATIVES À L'EMPLOI DU TEMPS TIRÉES DE L'ESG DE 1992

Introduction

Le présent article évalue la qualité des résultats des questions du recensement de 1996 qui portent sur les travaux ménagers non rémunérés et les soins non rémunérés aux enfants à partir de comparaisons entre ces résultats et les données de journaux personnels de l'emploi du temps tirées de l'Enquête sociale générale (ESG) de 1992. L'étude cherche à déterminer si certaines tendances d'écart entre les données de journaux et les données provenant de questions directes, relevées dans la documentation, sont observées dans des données raisonnablement comparables de certaines distributions choisies des résultats de l'ESG de 1992 et du recensement de 1996.

Seules les questions du recensement traitant des travaux ménagers et des soins aux enfants sont abordées dans cet article, les données de journaux de l'ESG de 1992 ne permettant pas d'analyser les heures consacrées aux soins aux personnes âgées. L'ESG de 1998 a tenté de corriger cette lacune grâce à des questions spéciales portant sur les soins dispensés aux personnes âgées.

Le présent article est le troisième et dernier rapport d'une série d'études sur l'évaluation de la qualité des questions du recensement relatives au travail non rémunéré. Le premier article de la série portait sur la qualité des données du recensement déterminée en fonction des taux de non-réponse et des niveaux d'imputation attribuables à des réponses manifestement erronées (Swain et Stone, 1998). Le deuxième article examinait les répercussions, sur le plan de la qualité, des écarts observés entre les distributions axées sur les données du recensement de 1996 et celles fondées sur les résultats de l'ESG de 1996, compte tenu du fait que les deux enquêtes utilisaient des questions identiques (Stone et Swain, 1999). Un quatrième article traitant des utilisations analytiques possibles des données du recensement est également disponible (Stone et Silver, 1998a).

Méthodologie

Compte tenu du nombre limité de comparaisons qu'il est possible d'effectuer entre les données provenant de journaux et les données provenant de questions directes (voir ci-dessous), l'hypothèse que l'on cherche à vérifier dans le cadre du présent article est fort simple. On s'attend généralement à ce que les modèles de distribution des estimations de l'emploi du temps axées sur la méthode du journal et des estimations fondées sur des questions directes soient raisonnablement similaires en ce qui a trait aux travaux ménagers. Dans le cas des soins aux enfants, la méthode des questions directes devrait produire des estimations beaucoup plus élevées que la méthode du journal, cette dernière ne tenant compte que des heures consacrées aux soins aux enfants à titre d'activité principale.

Encore une fois, en raison de la non-comparabilité fondamentale entre les données de journaux de l'ESG et les données du recensement, nous ne comparerons pas de chiffres précis. Nous cherchons plutôt à présenter les tendances des données au moyen d'une série de graphiques.

Comparabilité des données du recensement et de l'ESG

Bon nombre des sources de non-comparabilité entre l'ESG de 1996 et le recensement présentées dans le deuxième article s'appliquent également à l'ESG de 1992, notamment la méthode de collecte des données par interview téléphonique, les différences au chapitre de la période de référence, les déclarations par procuration et le traitement de la non-réponse. (Pour une discussion détaillée de ces différences, voir Stone et Swain, 1999.) Les données provenant de journaux de l'emploi du temps, toutefois, introduisent de nouvelles sources de non-comparabilité entre le recensement et l'ESG qui viennent s'ajouter aux différences précitées.

Comme nous l'avons mentionné dans le deuxième article, la période de référence constitue une source de non-comparabilité entre le recensement et l'ESG. À l'instar des données de l'ESG de 1996, les données de l'ESG de 1992 ont été recueillies sur une période d'un an, tandis que celles du recensement sont recueillies une seule fois à une date donnée. Outre la variabilité saisonnière des résultats, il convient de noter que les enquêtes ont été réalisées à des années différentes : 1992 dans le cas de l'ESG et 1996 dans le cas du recensement.

Une différence très importante et fondamentale entre les données du recensement et les données tirées des journaux de l'emploi du temps de l'ESG de 1992 réside dans les méthodes de collecte et d'estimation des renseignements sur les heures consacrées au travail non rémunéré. Dans le cas de données tirées de journaux, les personnes sont invitées à déclarer toutes les activités auxquelles elles se sont adonnées au cours d'une période de 24 heures (généralement la journée précédant l'interview) dans un journal de l'emploi du temps. Ces activités sont ensuite codées en fonction d'un ensemble prédéterminé de codes. Pour produire une estimation du nombre d'heures consacrées à une activité comme les travaux ménagers, on additionne les heures consacrées à diverses activités telles que la préparation des repas, la lessive et le ménage. C'est donc l'analyste qui détermine quelles activités constituent des travaux ménagers non rémunérés ou des soins non rémunérés.

Par contre, le recensement utilisait trois questions directes pour recueillir des renseignements sur les trois activités distinctes que regroupe le travail non rémunéré. Les répondantes et les répondants déclaraient un nombre estimé d'heures consacrées à chacune de ces activités au cours de la période de sept jours précédant le recensement. Pour réduire le fardeau de réponse, les questions du recensement étaient assorties de catégories de réponse permettant d'inscrire les heures consacrées à ces activités plutôt qu'un chiffre exact. Ce sont donc les répondantes et les répondants qui, en définitive, déterminaient quelles activités inclure dans chacune des questions portant sur le travail non rémunéré. Cette méthode peut donner lieu à de multiples possibilités d'interprétation par les répondantes et les répondants, particulièrement dans le cas de la question sur les travaux ménagers.

Bref, les sources de non-comparabilité entre les données du recensement et les données de journaux de l'emploi du temps de l'ESG sont si nombreuses qu'il convient d'adopter l'hypothèse de travail voulant que ces deux enquêtes mesurent, en fait, des variables différentes. Chacune des variables mesurées est utile pour une catégorie particulière de questions analytiques. Il se peut que ces différentes catégories de questions analytiques se chevauchent, mais elles ne sont pas équivalentes.

Méthodes de collecte de données : questions directes et journaux de l'emploi du temps

Dans sa recension des écrits sur la collecte des données axée sur la méthode du journal et sur les questions directes, Paillé (1994) observe que les estimations issues de la méthode du journal sont généralement jugées supérieures à celles fondées sur des questions directes sur le plan de l'exactitude. La méthode du journal s'appuie sur une période de référence plus courte (habituellement les 24 heures précédant l'enquête). Il est donc plus facile pour les répondantes et les répondants de se souvenir de leurs activités. Ces personnes inscrivent simplement leurs activités en ordre chronologique; celles-ci sont ensuite codées par l'analyste.

En revanche, dans la mesure où les journaux recueillent des renseignements visant une seule journée dans la vie des répondantes et des répondants, il est aisé de sous-estimer les activités peu courantes. Le travail bénévole est une activité à laquelle une répondante ou un répondant donné s'adonne souvent quelques jours par semaine ou par mois, et, par conséquent, les données tirées de journaux sous-estiment vraisemblablement la fréquence de cette activité dans la population (pour une discussion de cette question, voir Paillé, 1994).

Les journaux de l'ESG recueillent, en règle générale, des renseignements uniquement sur ce qu'on appelle les activités principales. Ils ne recueillent pas de données sur d'autres activités simultanées, notamment, et de façon fort importante, les soins aux enfants.

Les questions directes portent généralement sur une période de référence plus longue, par exemple, une semaine ou un mois, ce qui peut causer des difficultés d'ordre mnémonique pour les répondantes et les répondants. Ces personnes doivent également déterminer les activités qu'elles doivent inclure lorsqu'elles indiquent le nombre d'heures consacrées à certains travaux tels que les travaux ménagers⁶.

Enfin, comparativement aux données provenant de journaux, les données axées sur des questions directes peuvent donner lieu à une surestimation du nombre d'heures consacrées à des activités qui se déroulent souvent en même temps que d'autres. Cette surestimation possible est attribuable au fait que l'on demande aux répondantes et aux répondants d'estimer le nombre d'heures consacrées à chaque activité de façon distincte, de sorte que lorsqu'on fait la somme des heures déclarées à chacune des questions, le temps total estimé est souvent supérieur à la période de référence.

Cependant, cette caractéristique des données fondées sur des questions directes ne constitue pas une lacune lorsqu'il importe de tenir compte des tâches multiples accomplies au cours d'une période déterminée. Dans ce cas, il peut être plus facile de mesurer les activités souvent

faites simultanément au moyen de questions directes que par la méthode du journal de l'emploi du temps, qui restreint chaque moment de la journée à une seule activité principale. (Cette restriction n'est pas intrinsèquement liée aux données tirées de journaux de l'emploi du temps, mais elle s'applique aux journaux tenus dans le cadre de l'ESG.)

Dans ce contexte, les questions directes ne donnent pas lieu à une surestimation du nombre d'heures. Les analystes doivent simplement faire preuve de circonspection au moment de faire la somme du nombre d'heures consacrées à diverses activités. Il s'agit d'un problème mineur puisque les données fondées sur des questions directes sont rarement utilisées pour produire des estimations du nombre total d'heures consacrées au travail non rémunéré.

La question de la mesure des activités multiples au cours d'une période déterminée est fondamentale lorsqu'on cherche à estimer le nombre d'heures consacrées aux soins aux enfants. La plupart des spécialistes en matière de soins aux enfants sont d'avis qu'on ne peut estimer adéquatement le nombre total d'heures consacrées aux soins aux enfants que si l'on tient compte des activités simultanées.

Pour revenir aux hypothèses devant être vérifiées dans le présent article, la documentation tend à indiquer que les estimations du nombre d'heures consacrées aux travaux ménagers devraient être assez similaires pour les deux méthodes. En ce qui a trait aux heures consacrées aux soins aux enfants, toutefois, les questions directes devraient produire des estimations plus élevées (c'est-à-dire des proportions plus fortes dans les catégories supérieures de la distribution) que les données tirées des journaux de l'emploi du temps.

Travaux ménagers

La liste des activités qui, dans les données des journaux de l'ESG, servait à estimer le nombre d'heures consacrées aux travaux ménagers non rémunérés est tirée de la base de données du Système de comptes du travail total (SCTT) (Stone et Chicha, 1997). Il convient de noter que la définition du « travail non rémunéré pour soi-même et pour autrui » utilisée dans les comptes du travail total (CTT) est assez large, et il est fort improbable que les répondantes et les répondants au recensement auraient de leur propre chef inclus toutes les activités regroupées sous cette variable telle qu'elle est définie dans le SCTT. (Pour la définition complète du « travail non rémunéré pour soi-même et pour autrui » employée dans les CTT, voir l'annexe 3.A.) Par conséquent, les figures présentées ci-dessous offrent uniquement une indication approximative de la comparabilité des données sur les travaux ménagers non rémunéré provenant de journaux et de celles provenant de questions directes.

En ce qui concerne tout d'abord les données sur les femmes (figure 3.1), les distributions semblent relativement similaires, exception faite de la catégorie « de 2 à 4 heures ». Toutefois, ce résultat pourrait s'expliquer par les différences au chapitre de la liste d'activités qu'il est possible de réunir sous la rubrique des travaux ménagers, comme on l'a mentionné précédemment.

Quant aux données sur les hommes, les distributions établies à partir des deux sources semblent assez similaires. Les données tirées de journaux indiquent une proportion plus

élevée d'hommes consacrant de 4,3 à 8,5 heures aux travaux ménagers non rémunérés que les données du recensement. Là encore, ces écarts peuvent être attribuables aux différentes définitions des travaux ménagers.

Par contre, la proportion nettement plus élevée d'hommes ne déclarant aucune heure consacrée aux travaux ménagers dans le recensement par rapport à l'ESG pourrait nous porter à croire que cette activité est sous-estimée pour les hommes dans les données du recensement. Bien que plusieurs raisons puissent expliquer cette situation, la raison la plus souvent évoquée est la déclaration par procuration. On doit alors présumer que la plupart des déclarations par procuration sont faites par les femmes et que celles-ci sous-estiment systématiquement les travaux ménagers accomplis par leur conjoint. Il se peut, cependant, que les hommes eux-mêmes sous-estiment le temps qu'ils consacrent aux travaux ménagers, soit parce qu'ils estiment que cette question du recensement ne s'applique pas à eux ou parce qu'ils jugent que ces tâches ou les heures qu'ils y consacrent sont trop peu importantes pour en tenir compte. Par exemple, le fait de sortir les ordures ou de changer une ampoule prend peu de temps, et ces tâches peuvent être ignorées ou même oubliées durant la semaine.

Soins aux enfants

La figure 3.2 compare le nombre d'heures consacrées aux soins aux enfants parmi les répondantes et les répondants vivant avec une enfant ou un enfant âgé de moins de 15 ans. Dans ce cas-ci, les résultats obtenus confirment clairement notre hypothèse initiale. Chez les hommes comme chez les femmes, la question directe du recensement donne lieu à des estimations plus élevées du temps consacré à cette activité comparativement aux données de l'ESG tirées de journaux de l'emploi du temps.

Pour ce qui est des distributions visant les femmes, la figure 3.2 montre, pour l'ESG, une forte proportion de réponses dans la catégorie des « 0,7 à 2,0 heures ». Ainsi, les données de l'ESG indiquent que seulement 20 % des femmes environ consacrent plus de deux heures par jour aux soins aux enfants. S'il est possible que les enfants âgés de plus de cinq ans ne nécessitent pas plus de deux heures de soins pendant lesquelles l'attention de la mère porte entièrement sur l'enfant, ces résultats n'ont de sens que si l'on cherche à mesurer les soins aux enfants à titre d'activité principale. Mais comme nous l'avons soutenu, une mesure véritable des heures consacrées aux soins aux enfants doit également tenir compte des activités secondaires ou simultanées. Par conséquent, en dépit des lacunes des données du recensement sur les soins aux enfants que nous avons décrises dans notre deuxième article, la distribution axée sur les données du recensement est vraisemblablement plus utile à l'étude du comportement des enfants que celle fondée sur les données de journaux de l'ESG.

La figure 3.3 tend à confirmer cette critique des données de journaux de l'ESG. Nous nous intéressons ici aux ménages comptant au moins une enfant ou un enfant âgé de moins de six ans. Nous nous attendons donc à ce que cette catégorie de personnes obtienne le nombre le plus élevé d'heures allouées aux soins aux enfants. Outre les variables du recensement et de l'ESG se rapportant aux soins aux enfants, la variable « travaux relatifs aux enfants » des CTT est prise en considération. Dans les CTT, on élargit la définition des soins aux enfants de l'ESG pour créer une nouvelle variable appelée « travaux relatifs aux enfants » qui

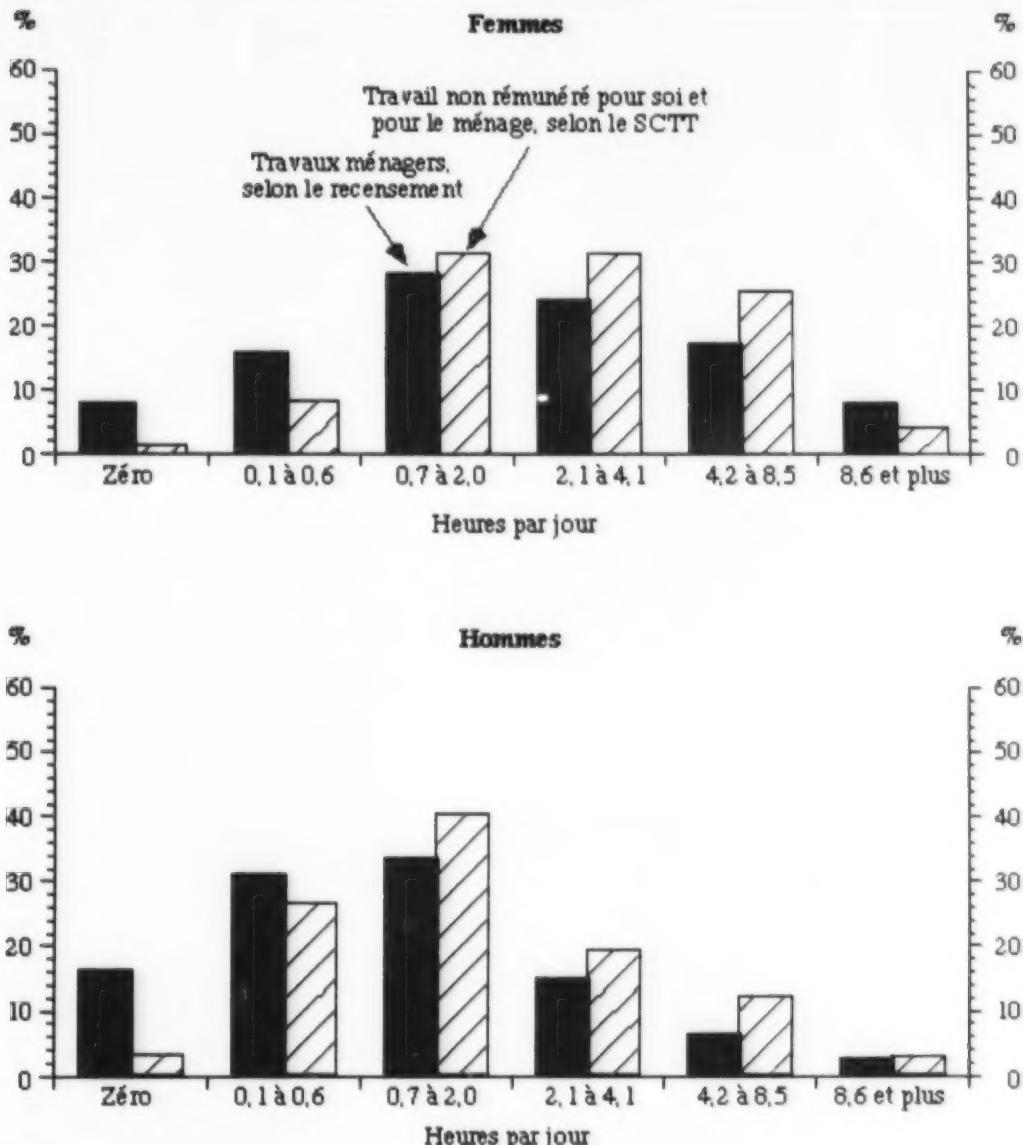
englobe plusieurs activités supplémentaires comme les travaux ménagers et les courses si l'enfant est présent à ce moment-là (voir la définition à l'annexe 3.A).

Le fait d'élargir ainsi cette définition a pour effet de déplacer la distribution de l'ESG vers la droite, ce qui suggère une somme de travail plus considérable que ce qu'indiquent les données axées uniquement sur l'activité principale. Quoi qu'il en soit, même la distribution s'appuyant sur la définition des CTT demeure bien en deçà de la distribution fondée sur les données du recensement, particulièrement pour ce qui est de la catégorie supérieure des « 8,6 heures ou plus ». Si, d'après les données du recensement, on semble surestimer ces activités, il convient de noter que, dans notre deuxième étude, nous avons observé que la proportion de femmes ayant déclaré 60 heures ou plus au chapitre des soins aux enfants dans le cadre du recensement est systématiquement inférieure à la proportion obtenue au moyen de la question directe de l'ESG de 1996.

Comme nous l'avons mentionné précédemment, dans le cas des hommes également, l'estimation du temps consacré aux soins aux enfants établie à partir de la question directe du recensement est supérieure à celle établie à partir de la méthode du journal.

Il importe de souligner l'écart considérable observé entre le recensement et l'ESG quant à la proportion d'hommes ayant déclaré n'avoir consacré aucune heure aux soins aux enfants. La proportion des hommes obtenue à partir des données de l'ESG correspond pratiquement au double de la proportion des hommes établie à partir des données du recensement pour cette catégorie. Ces résultats pourraient s'expliquer par le fait que les données de journaux considèrent les soins dispensés aux enfants à titre d'activité principale.

Graphique 3.1. Répartition de la population par catégories d'heures consacrées aux travaux ménagers (recensement de 1996) et d'heures consacrées au travail non rémunéré pour soi et pour le ménage (SCTT 1992)¹
 (Personnes de 15 ans et plus)



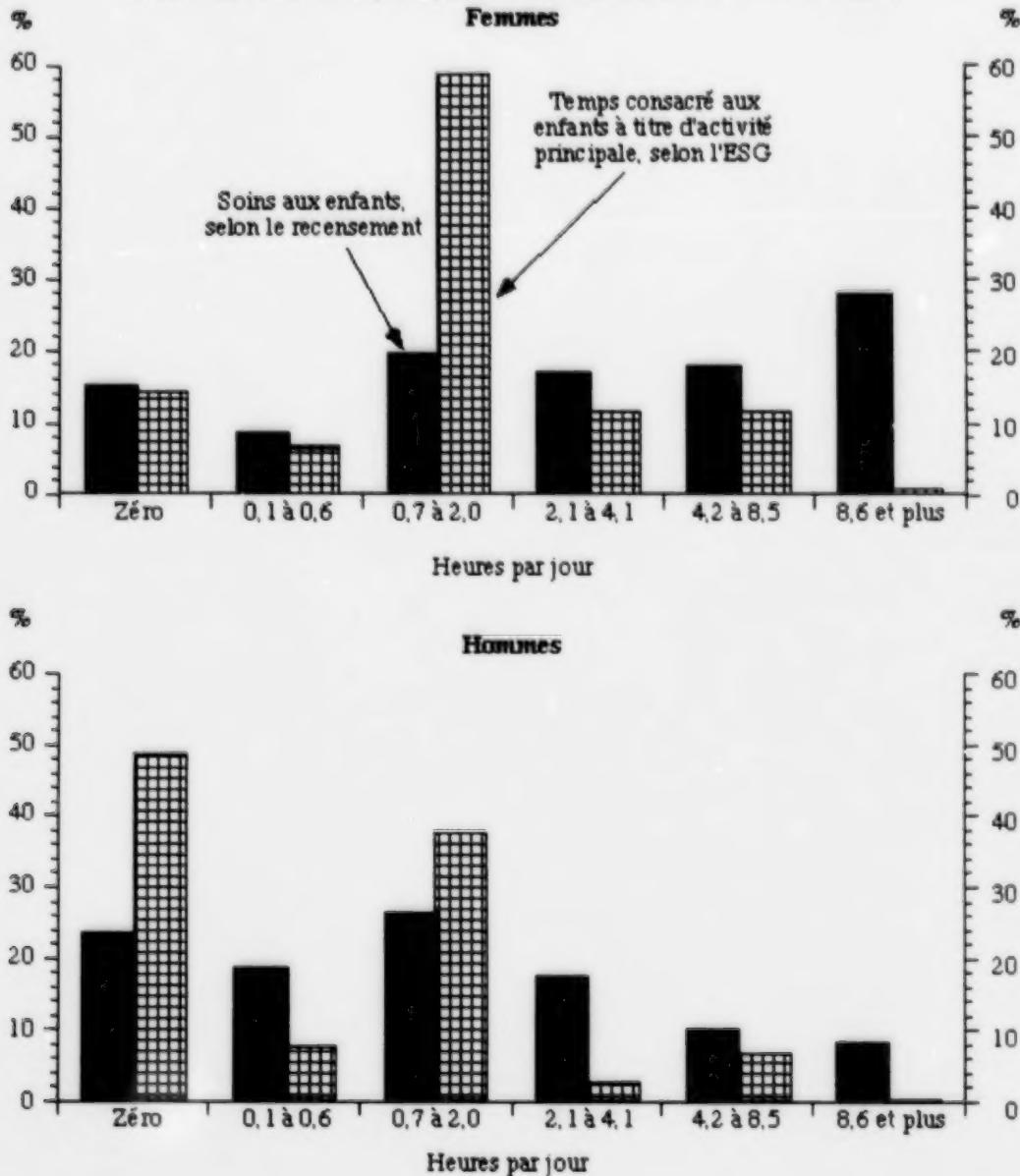
Note :

¹ Système de comptes du travail total, d'après l'Enquête sociale générale de 1992.

Source :

Statistique Canada, Recensement de 1996 et Système de comptes du travail total (d'après l'Enquête sociale générale).

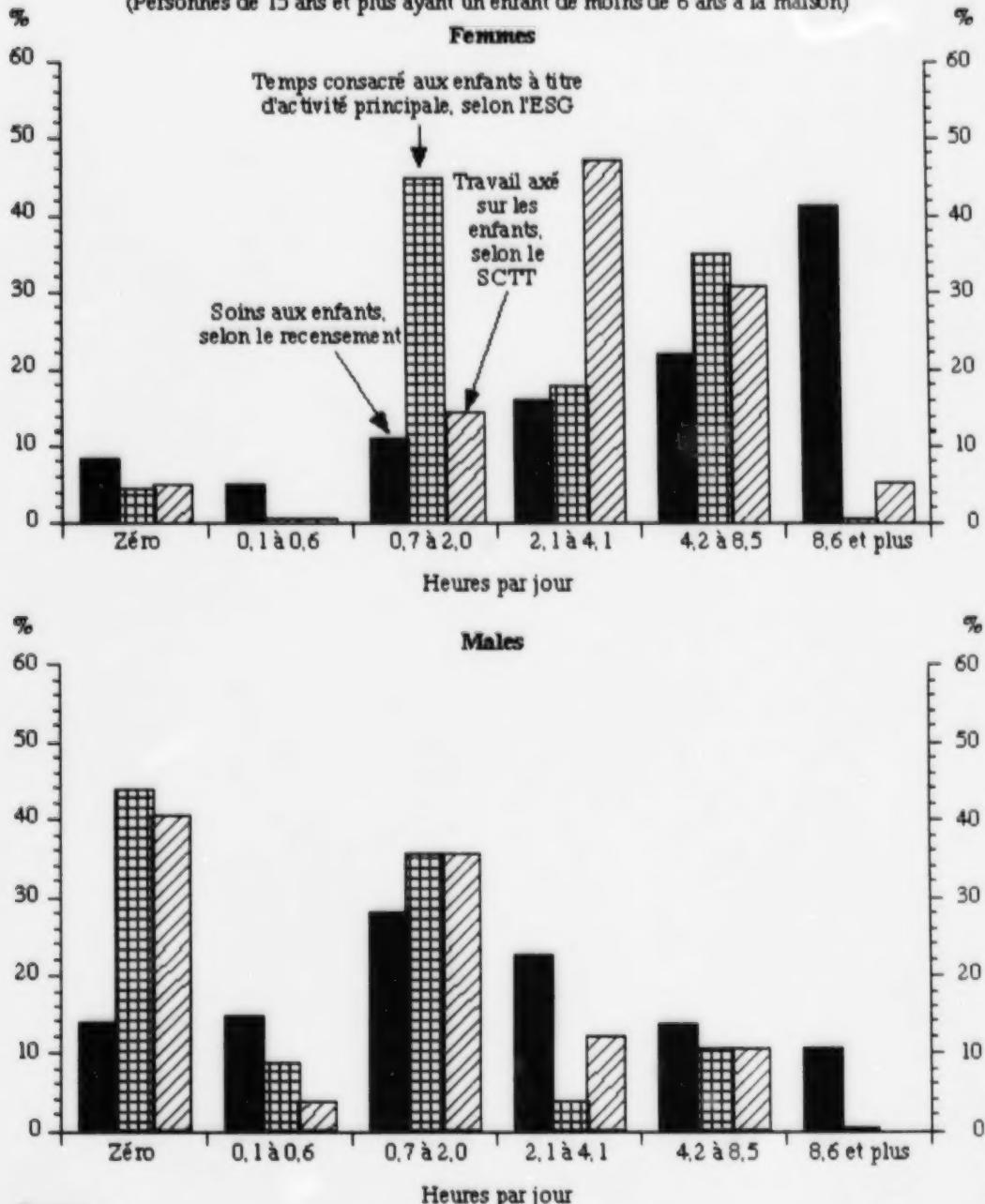
**Graphique 3.2. Répartition de la population par catégories d'heures consacrées aux soins non rémunérés aux enfants, selon le sexe - Recensement de 1996 et
Enquête sociale générale de 1992**
(Personnes de 15 ans et plus ayant un enfant de moins de 15 ans à la maison)



Source :
Statistiques Canada, Recensement de 1996 et Enquête sociale générale de 1992.

Graphique 3.3. Répartition de la population par catégories d'heures consacrées aux soins non rémunérés aux enfants, selon le sexe – Recensement de 1996 et Enquête sociale générale de 1992

(Personnes de 15 ans et plus ayant un enfant de moins de 6 ans à la maison)



Source :

Statistique Canada, Recensement de 1996, Enquête sociale générale de 1992 et Système de comptes du travail total (d'après l'Enquête sociale générale).

Conclusion

En conclusion, nous sommes d'avis qu'il n'est pas possible, dans le cadre de cette évaluation des données du recensement, de soutenir que les données provenant de journaux de l'emploi du temps sont globalement plus exactes que les données obtenues au moyen de questions directes.

Puisque le recensement mesure seulement trois aspects particuliers du travail non rémunéré, on ne peut calculer le volume total ou global du travail non rémunéré à l'aide des données du recensement. On doit plutôt étudier le temps alloué aux travaux ménagers non rémunérés, aux soins non rémunérés aux enfants et aux soins non rémunérés aux personnes âgées comme des ensembles d'activités distincts et non cumulatifs.

L'hypothèse que nous avons adoptée dans le présent article est la suivante : la comparaison des données obtenues selon la méthode du journal de l'emploi du temps et selon la méthode des questions directes devrait indiquer des tendances similaires de distribution pour les travaux ménagers, mais les questions directes devraient produire des estimations nettement plus élevées que les données de journaux dans le cas des soins prodigués aux enfants. Les résultats présentés dans le présent article confirment cette hypothèse. Même les tendances relatives aux travaux ménagers, bien qu'elles se fondent sur des définitions très différentes (selon les CTT et ce qu'on s'attend à obtenir dans le cadre du recensement), n'indiquent pas que les données du recensement divergent de celles provenant des journaux de l'emploi du temps.

Notre analyse des données sur les soins aux enfants révèle que les données provenant de journaux ne tiennent pas compte de la situation dans son ensemble. Le concept des tâches multiples est essentiel à la compréhension du temps consacré aux soins aux enfants. Nous estimons, par conséquent, que les données du recensement sont supérieures à celles tirées des journaux de l'emploi du temps pour mesurer cette activité.

Enfin, en ce qui concerne les soins non rémunérés dispensés aux personnes âgées, les données des journaux de l'emploi du temps ne permettent d'effectuer aucune analyse raisonnable de cette activité.

Annexe 3.a : les définitions opérationnelles du travail ayant une valeur économique

« [La présente annexe] est un ensemble de définitions opérationnelles qui sont strictement reliées aux propriétés et aux limitations de l'ESG de 1992. Les définitions représentent souvent un effort pour imputer les bénéficiaires spécifiques des activités de travail selon la nature de l'activité et la présence de personnes données pendant la poursuite de l'activité. »

« Cette imputation était nécessaire parce que le questionnaire de l'ESG ne demandait pas aux répondants d'identifier les personnes perçues comme les bénéficiaires de leurs activités de travail. Cet écart est une limitation majeure de la base de données de l'ESG, dans le contexte du SCTT. »

[...]

« L'utilisation d'un langage usuel ne permet pas de rendre compte clairement de certaines des définitions ci-dessous, et cela, à cause de la nécessité d'utiliser des structures de logique [booléenne] complexes pour la construction de certaines définitions. Par conséquent, l'ensemble des définitions suivantes composées de termes familiers sont moins précises que le code correspondant du logiciel. »

[...]

« *Destination 6 : L'enfant* – [...] accomplir du travail domestique, ou faire les courses ou prodiguer des soins à l'enfant, et ce, en contact avec l'enfant pendant l'accomplissement du travail. »

[...]

« *Destination 9 : Soi-même et les autres membres du ménage* – [...] préparation de repas, vaisselle, emplettes, lavage, habillage, rangement, soins médicaux à domicile, aide et soins personnels aux adultes, et ce, en contact avec personne pendant le déroulement du travail (imputation de produit du travail de valeur économique — services que vous achèteriez si vous ne pouviez pas le faire vous-même). »

Source : Stone et Chicha. (1996).

RÉFÉRENCES

Paillé, B. (1994). *Estimation du temps consacré aux activités productives non rémunérées au Canada, 1992 : évaluation des données de l'Enquête sociale générale*. Document de travail n° 10 de l'Enquête sociale générale, Statistique Canada.

Stone, L.O. (1999). « Computing a Household-Level Index for Time Spent Doing Child Care Work. » Document non publié, 2 pages.

Stone, L.O. et M.-T. Chicha. (Octobre 1997). *Constructing the Statistics Canada Total Work Accounts System – Principles and Problems in Allocating Unpaid Work Outputs Among Alternative Destinations Using Conventional Time Use Data*. Article présenté dans le cadre de la 19^e réunion de l'International Association for Time Use Research, Stockholm, Suède.

—. (1996). *Le système de comptes du travail total de Statistique Canada*. N° 89-549-XPF au catalogue de Statistique Canada. Ottawa, ministre de l'Industrie.

Stone, L.O. et C. Silver. (1998a). « Potential Contributions of the 1996 Census to Policy Applications of Data on Unpaid Work. » Document non publié.

—. (1998b). *The Development of a Measure of a Household's Total Time Spent Caring for Children Using the 1996 Census Data*. Document de travail.

Stone, L.O. et S. Swain. (1999). *Quality of the 1996 Census Unpaid Work Data in the Light of Comparisons with 1996 GSS Data Based on Identical Questions*. Document de travail.

Swain, S. et L. Stone. (1998). *Evaluation of the 1996 Census Questions on Unpaid Work – Report No. 1 Non-Response and Multiple Response Rate*. Document de travail.

GLOSSAIRE

Enfants

Fils et/ou filles jamais mariés de couples actuellement mariés, de couples vivant en union libre ou de parents seuls.

Famille de recensement

Couple actuellement marié (avec ou sans fils et/ou filles jamais mariés des deux conjoints ou de l'un d'eux), couple vivant en union libre (avec ou sans fils et/ou filles jamais mariés des deux partenaires ou de l'un d'eux) ou parent seul (peu importe son état matrimonial) demeurant avec au moins un fils ou une fille jamais marié.

Famille économique

Groupe de deux personnes ou plus qui vivent dans le même logement et qui sont apparentées par le sang, par alliance, par union libre ou par adoption.

Logement

Ensemble de pièces d'habitation qu'une personne ou un groupe de personnes habite ou pourrait habiter.

Logement collectif

Établissement commercial, institutionnel ou communautaire. Sont inclus dans cette catégorie les pensions et maisons de chambres, les hôtels, motels et maisons de chambres pour touristes, les maisons de repos, les hôpitaux, les résidences de personnel, les casernes (camps militaires), les camps de chantier, les prisons, les centres d'accueil, les foyers collectifs, etc.

Logement privé

Ensemble distinct de pièces d'habitation ayant une entrée privée donnant sur l'extérieur ou sur un corridor, un hall, un vestibule ou un escalier commun à l'intérieur.

Présence d'enfants

Variable permettant de classer les personnes âgées de 15 ans et plus vivant dans une famille de recensement en deux catégories : celles sans enfant à la maison et celles avec enfants. Les personnes vivant avec des enfants sont également regroupées selon le groupe d'âge de l'ensemble de leurs enfants.

NOTES

¹ Aujourd'hui, les analystes de données reconnaissent généralement que, pour ce qui est des soins aux enfants à tout le moins, les catégories clés de réponse du recensement de 1996 (voir l'annexe A) sont nettement trop larges et devront être mieux circonscrites si l'on prévoit recueillir des données de cette nature dans le cadre de recensements ultérieurs.

² Le terme « questionnaire 2B » fait référence au questionnaire comportant le plus de questions et destiné à être rempli par 20 % seulement de l'ensemble des répondantes et des répondants au recensement. Si le questionnaire 2B était constitué le principal instrument de collecte des données du recensement, le « questionnaire 2D du recenseur » a été utilisé dans les régions éloignées et du Nord. Ce questionnaire se distingue de la formule 2B du fait que certains des exemples cités dans les questions sur le travail non rémunéré ont été modifiés pour qu'ils soient adaptés aux conditions de vie du Nord. En outre, ce questionnaire était rempli par une intervieweuse ou un intervieweur, tandis que le questionnaire 2B était rempli par les répondantes et les répondants. Moins de 2 % des ménages du Canada ont reçu le questionnaire 2D.

³ Comme nous l'avons mentionné précédemment, le questionnaire 2B (défini à la note 1) est distribué à un ménage sur cinq au Canada. Cet échantillon de répondantes et de répondants est pondéré de façon à permettre la production d'estimations visant l'ensemble de la population canadienne. Les taux non pondérés fournissent des estimations visant uniquement les répondantes et les répondants ayant reçu le questionnaire 2B.

⁴ Afin de réduire le temps et les fonds alloués à la recherche d'enregistrements donneurs, les enregistrements des répondantes et des répondants sont normalement répartis en strates qui constituent des critères d'appariement définis au sens large, p. ex. les femmes âgées de 25 à 44 ans. Les enregistrements comportant des erreurs et les enregistrements donneurs potentiels doivent provenir de la même strate; l'enregistrement donneur et l'enregistrement comportant des erreurs ne peuvent provenir de strates différentes. Outre les restrictions de recherche imposées par la strate, on impose une restriction physique, la base de données du recensement étant subdivisée en cinq bases plus restreintes pour faciliter le traitement. Ces cinq bases sont les suivantes : ménages de l'Est, du Québec, de l'Ontario, de l'Ouest et d'outre-mer. Par conséquent, l'enregistrement donneur et l'enregistrement comportant des erreurs seront appariés à tout le moins en fonction de critères généraux définis par la strate et l'emplacement géographique.

⁵ Ces chiffres excluent les résultats relatifs aux ménages d'outre-mer. Il s'agit là d'une base de données de taille réduite regroupant les renseignements visant un groupe varié de répondantes et de répondants œuvrant dans les services diplomatiques, militaires ou travaillant sur des navires. Par conséquent, le taux d'appariement entre des enregistrements donneurs parfaitement assortis pour ces personnes est plus faible; il s'établit à 73 % dans le

cas des variables relatives aux travaux ménagers et aux soins aux enfants et à 90 % dans le cas de la variable relative aux soins aux personnes âgées. Toutefois, il a été possible de trouver un enregistrement donneur pour tous les enregistrements comportant des erreurs. Le potentiel qu'ont les enregistrements des ménages d'outre-mer de modifier les résultats dans le cadre de presque toutes les applications analytiques des données reste minime.

* Le fait que l'une des sources s'appuie sur un ensemble déterminé d'activités regroupées sous la rubrique « travaux ménagers » tandis qu'une autre se fonde sur un ensemble non défini d'activités signifie probablement que les estimations axées sur ces sources sont fondamentalement non comparables.